BACHELOR'S DEGREE THESIS

Degree in Mathematics

Title: Simulation and Estimation of Lévy driven stochastic processes.

Author: Pere Miquel Brull Borràs

Advisor: Argimiro Arratia Quesada

Department: Ciències de la Computació

Academic year: 2016-2017

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Universitat Politècnica de Catalunya
Facultat de Matemàtiques i Estadística

Degree in Mathematics

Bachelor's Degree Thesis

Simulation and Estimation of Lévy driven stochastic processes

Pere Miquel Brull Borràs

Supervised by Argimiro Arratia Quesada
Department of Computer Science & BGSMath

June, 2017

Thanks to professor A. Arratia for his time and patience and all his work behind this project.

To Sara.

Abstract

We carry on an exploration of Lévy processes, focusing on instrumental definitions that ease our way towards their simulation. Beginning with the basic Brownian motion and Poisson processes, we then explore algorithms to simulate Lévy processes, do maximum likelihood estimations, and follow this exploratory road studying a maximum likelihood methodology to estimate the parameters of a one dimensional stationary process of Ornstein Uhlenbeck type that is constructed via a self-decomposable distribution D. Finally, we also present the maximum empirical likelihood method specifically for Lévy processes as an alternative to the classical maximum likelihood estimation methodology, when the density function is unknown.

Keywords

Lévy process, Ornstein-Uhlenbeck processes, likelihood inference, process estimation, empirical likelihood.

1. Introduction to Lévy processes

1.1 Basic Lévy Processes

A stochastic process $\mathbf{X} = \{X(t)\}$ is a collection of random variables uniquely associated with an element of a set, usually $t \in \mathbb{N}$ which we will refer as time, and is usually denoted as X_t. Each random variable of the collection takes values from the same state space. As a stochastic process is a random process it can show many different outcomes, called sample functions or realizations.

One of the most basic stochastic processes are the so-called Brownian Motions or Wiener processes, and these are one of the best known Lévy Processes.

Definition 1.1. A *Wiener process* or *Brownian Motion* $\mathbf{W} = \{W_t, t > 0\}$ is a stochastic process characterized by the following properties:

1. $W_0 = 0$ a.s.
2. W has independent increments: $W_{t+u} - W_t$ is independent of $\sigma\ (W_s : s \leq t)$ for $u \geq 0$.
3. W has Gaussian increments: $W_{t+u} - W_t$ is normally distributed, $W_{t+u} - W_t \sim \mathcal{N}(0, u)$.
4. W has continuous paths: With probability 1, W_t is continuous in t.

We can also characterize those processes with the *Lévy characterization*: a Wiener process is a continuous martingale a.s with $W_0 = 0$, where a martingale is nothing but a fair game, meaning that if we know the filtration until time s, the expectancy will not go up nor down at future times, i.e. $\mathbb{E}[W_t|\mathcal{F}_s] = W_s$, $\forall t \geq s$, and $\mathbb{E}[|X_t|] < \infty$, $\forall t$.

The simulation of these processes is based on the normality of their increments. Thus, we need to prepare a grid on the time points $\{n\Delta t, n = 0, 1, ...\}$ and follow this recursive scheme:

$$W_0 = 0, \quad W_{n\Delta t} = W_{(n-1)\Delta t} + \nu_n\sqrt{\Delta t} \tag{1}$$

where $\{\nu_n, n = 1, 2, ...\}$ is a series of standard Normal random numbers. The simulation algorithm would have these steps:

1. Choose the time interval, usually $[0, T]$, and the number of the grid intervals N.
2. Set $\Delta t = T/N$.
3. Generate N standard Normal random numbers.
4. Apply the scheme (1).

To asses if simulation technique we are following is reliable, meaning that with enough time points the result truly converges to a well sampled Brownian Motion, one should not rely on graphics but on statistics. Although this case is fairly straight forward, we can compute the sample mean and variance of the points we generate and check if they coincide with the theoretical moments. As we just said, the simulation is based on the increments, so we expect the mean and variance to be 0 and Δt respectively.

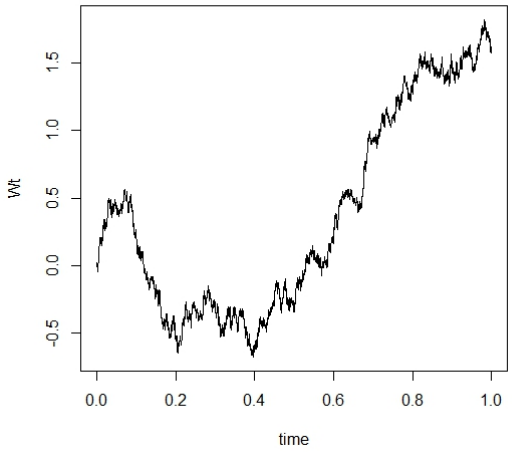

Figure 1: Simulation of a Brownian Motion sample path with $N = 2^{12}$ time intervals.

As this is a method with linear order on the number of intervals in the grid: $\mathcal{O}(N)$, we can sample various processes really fast to check their moments. Sampling 1000 Brownian Motions, we obtain the following Root Mean Square Errors of their sample moments against theoretical moments:

$$\textit{Mean} : \textit{RMSE} = 0.0014, \quad \textit{Variance} : \textit{RMSE} = 4.523e-05.$$

Definition 1.2. According to Sato [16], a stochastic process $\{X_t : t \geq 0\}$ on $\mathbb{R}$ is a *Poisson process* with parameter $c > 0$ if it is a Lévy process and, for $t > 0$, X_t has Poisson distribution with constant mean. This would be an axiomatic definition of the process, making it just a particular case of a Lévy process, which we will define formally in section 1.2 below.

However, we will focus on another characterization of this family which is a bit more practical, introduced by Stirzaker in [18], where you can also find all the proofs of the results we are going to present. Following his introduction, people often face an unending random sequence of events, such as customers entering a shop at random times and, occasionally making random purchases, or emails arriving at unpredictable times with an unknown percentage of junk emails. For the modelling of those sequences we need a *counting process*.

Definition 1.3. $N(t),\ t \geq 0$ is a counting process if it is non-negative, integer-valued and non-decreasing in t. Then, $N(t) - N(s)$ is the number counted in $(s, t]$.

From all counting processes, the most important one would be the *Poisson process*:

Definition 1.4. The counting process $N(t)$ is a *Poisson process* with intensity λ if $N(0) = 0$, and

1. $N(t-s) - N(s)$ is a Poisson$(\lambda t), 0 < s < s + t$.
2. $N(t)$ has independent increments: For any $0 \leq t_0 < t_1 < ... < t_n < \infty, N(t_0), N(t_1) - N(t_0), ...,$ $N(t_n) - N(t_{n-1})$ are independent.

We also state these results that are important for the construction of Poisson processes:

Proposition 1.5. *Let $\{X_r; r \geq 1\}$ be independent and exponential(λ); if we define $T_n = \sum_{r=1}^{n} X_r$, with $T_0 = 0$ and set*

$$N(t) = \max\{n : T_n \geq t\}.$$

Then T_n is the time arrival of the nth event, and $N(t)$ is the number that have arrived by time t. Moreover, $N(t)$ is a Poisson process, and the converse result can also be proved, being that if $N(t)$ is a Poisson process, then the arrival times are independent and exponential.

Finally, this last result is the one that greatly simplifies the simulation of the process with the **uniform method**:

Theorem 1.6. *Let $N(t)$ be a Poisson process. Conditional on $N(t) = m$, the positions of these m events in $[0, t]$ are independently and uniformly distributed on $[0, t]$.*

With these ideas we have a process defined by random unitary jumps in the case of the Poisson process, and jumps set by an exponential distribution regarding the compound Poisson process.

Thus, as stated in [17], if we need to simulate a Poisson process with intensity parameter $\lambda > 0$ up to a time point $T > 0$, we can first generate a random number N which is Poisson(λT) distributed, or just set it to the number of desired jumps in the interval, but note that then the intensity λ will be unknown. Then, we simulate N independent random uniform numbers $u_1, \ldots, u_N$ and order the sequence to obtain the time points $Tu_{(1)}, \ldots, Tu_{(N)}$ of the jumps. The scheme would be as follows:

1. Generate $N \sim$ *Poisson*(λT) or set it to a fixed desired number.
2. Simulate N random uniform numbers.
3. Sort them to obtain the jump times.

This scheme allows us to simulate a Poisson process in $\mathcal{O}(m)$ steps, where m is the total number of jumps (events) preassigned to happen. However, we are missing the time discretization in that interval, and we may want the Poisson process path defined in the time grid. For that, we found another method also proposed by Schoutens in [17], the **method of the exponential spacings**:
Given a Poisson process $\mathbf{N} = \{N_t, t \geq 0\}$ with intensity parameter λ, we will use the fact that the inter-arrival times of the jumps of this process follow an exponential distribution with mean λ^{-1}, as the exponential is the paradigm of a distribution holding the Markov property of being memoryless. One can obtain these values from a Uniform(0,1) random number u_n, by

$$e_n = -\log(u_n)/\lambda. \tag{2}$$

Then, let

$$s_0 = 0, \quad s_n = s_{n-1} + e_n, \ n = 1, 2, , \ldots, N \tag{3}$$

where N+1 is the number of points in the time grid, so $\Delta_t = \Delta = T/N$. Now we can sample a path of the Poisson process N in the time points $n\Delta$, $n = 1, 2, \ldots, N$:

$$N_0 = 0, \quad N_{n\Delta} = \sup(k : s_k \leq n\Delta) \tag{4}$$

Again, these are the steps to follow:

1. Generate N exponential random variables or

2. Generate N uniform random values and transform to the exponential using (2).
3. Create the vector s following (3).
4. Finally obtain the whole sample path with the formula (4).

This also yields a linear simulation of a Poisson process of order $\mathcal{O}(n)$, where n is the number of points in the discretization, and it doesn't just gives us the jump times, but the whole sample path of the process.

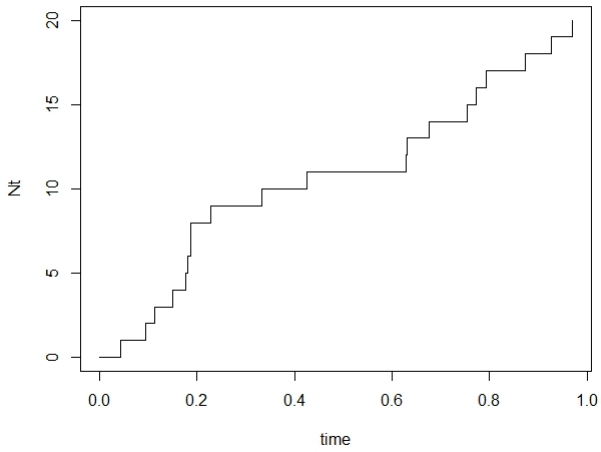

Figure 2: Simulation of Poisson process with the uniform method for $N = 20$

1.2 The Lévy process

Definition 1.7. A Lévy process $\mathbf{Z} = \{Z_t, t > 0\}$ is a *càdlàg* process (*"continue à droit, limite à gauche"*) that shares the first two properties with the Wiener processes, as it holds that $Z_0 = 0$ and has independent increments, but now we allow it to have jumps and we do not impose any fixed distribution, so we rewrite the last points:

3. Z has stationary increments: $\forall s < t$, Z_t - Z_s is equal in distribution to Z_{t-s}.
4. Z is continuous in probability: $\forall \epsilon > 0$ and $t \geq 0$, it holds that

$$\lim_{h \to 0} \mathbb{P}(|Z_{t+h} - Z_t| > \epsilon) = 0.$$

Before we continue with the simulations, we will state some important results about Lévy processes, most of them extracted from the work of Sato in [16].

Definition 1.8. A probability distribution is *infinitely divisible* if it can be expressed as the probability distribution of the sum of an arbitrary number of independent and identically distributed random variables. Another definition of being infinitely divisible is: A probability measure μ on $\mathbb{R}^d$ is infinitely divisible if, for any positive integer n, there is a probability measure μ_n on $\mathbb{R}^d$ such that $\mu = \mu_n^n$, where this denotes the n-fold convolution of the probability measure μ with itself, that is,

$$\mu^n = \mu^{n*} = \underbrace{\mu * ... * \mu}_{n}.$$

Since the convolution can be seen as expressed by the product in characteristic functions, we say that μ is infinitely divisible if and only if, for each n, an nth root of its characteristic function can be chosen in such a way that is the characteristic function of some probability measure.

The concept of being infinitely divisible is of major importance. There is a one-to-one correspondence between infinitely divisible distributions and the collection of all Lévy processes in the following theorem:

Theorem 1.9. *i If $\{X_t : t \geq 0\}$ is a Lévy process in law on $\mathbb{R}^d$, then for any $t \geq 0, P_{X_t}$ is infinitely divisible and, letting $P_{X_1} = \mu$, we have $P_{X_t} = \mu^t$, where as μ is an infinitely divisible distribution, for every $t \in [0, \infty)$, μ^t is definable and infinitely divisible.*

ii Conversely, if μ is an infinitely divisible distribution on $\mathbb{R}^d$, then there is a Lévy process in law $\{X_t : t \geq 0\}$ such that $P_{X_1} = \mu$.

iii If $\{X_t\}$ and $\{X_t'\}$ are Lévy processes in law on $\mathbb{R}^d$ such that $P_{X_1} = P_{X_1'}$, then $\{X_t\}$ and $\{X_t'\}$ are identical in law.

Before advancing to the next important result we are going to go through some examples to better understand and visualize this fundamental concept by the form of their characteristic functions. For a distribution μ, we will denote as $\hat{\mu}(z)$ its characteristic function. As we will present later on, this notation comes from the application of the Fourier transform.

Example 1.10.

- **Gaussian**: We call μ a Gaussian distribution if

$$\hat{\mu}(z) = e^{\gamma i z - \frac{\sigma^2 z^2}{2}}.$$

Then, by choosing the parameters appropriately we can get the nth root of these distributions. This result also holds on $\mathbb{R}^d$, where z and γ are vectors and A is a nonnegative-definite symmetric matrix.

$$\hat{\mu}(z) = \exp(-\frac{1}{1}\langle z, Az\rangle + i\langle \gamma, z\rangle).$$

- **Gamma Process**: Which has the following characteristic function

$$\hat{\mu}(z) = (1 - \frac{iz}{b})^{-a}.$$

Again, by tweaking the parameters we obtain the root of the same distribution.

- **Tempered Stable Process**: Whose characteristic function is

$$\hat{\mu}(z) = \exp(ab - a(b^{1/\kappa} - 2iz)^{\kappa}).$$

- **The Inverse Gaussian Process**: This process is followed by the random time at which a standard Brownian Motion with drift $b > 0$ first reaches the positive level $a > 0$. Its characteristic function is

$$\hat{\mu}(z) = \exp(-a(\sqrt{-2iz + b^2} - b)).$$

- Many more process among which we have the Simple and Compound Poisson processes, the Cauchy process, the Generalized Inverse Gaussian process and the Variance Gamma process.

  However, there are distributions whose infinite divisibility is more difficult to prove and have dedicated papers to that task, as the Student's *t*-distribution, Pareto distribution, F-distribution and logistic distribution.

Theorem 1.11. *(Lévy-Khintchine representation) The distribution of a Lévy process is characterized by it characteristic function, which is given by the Lévy-Khintchine formula. If X is a Lévy process, then its characteristic function at time $t \geq 0$ is:*

$$\phi_X(\theta)(t) := \mathbb{E}\Big[e^{i\theta X(t)}\Big] = \exp\left(t(ai\theta - \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}\setminus\{0\}} (e^{i\theta x} - 1 - i\theta x\mathbf{I}_{|x|<1})\,\Pi(dx))\right),$$

here $a \in \mathbb{R}$, $\sigma \geq 0$, $\mathbf{I}$ is the indicator function and Π is a sigma-finite measure called the Lévy measure of X, satisfying the property:

$$\int_{\mathbb{R}\setminus\{0\}} 1 \wedge x^2 \Pi(dx) < \infty,$$

and (a, σ^2, Π) form the Lévy generating triplet. The representation with a given Lévy triplet is unique. Conversely, given a Lévy triplet with a measure satisfying the above property, then there exists an infinite divisible distribution whose characteristic function is given by the Lévy-Khintchine formula.

Thanks to the previous theorem, this formula does not only gives a representation of characteristic functions for Lévy process, but for all infinitely divisible distributions. An extension of infinitely divisibility, which will be of use later on, is the concept of self-decomposability.

Definition 1.12. A distribution μ is said to be self-decomposable if for any $0 < a < 1$, there exists a distribution υ_a such that

$$\phi_\mu(\vartheta) = \phi_\mu(a\vartheta)\phi_{\upsilon_a}(\vartheta).$$

Alternatively, a random variable X is said to have a self-decomposable distribution if for any $0 < a < 1$, there exists a random variable Y_a, independent of X, such that

$$X \overset{d}{=} aX + Y_a.$$

Any non-degenerate distribution μ (*i.e.*, it takes more than one value) which is also self-decomposable, is absolutely continuous and infinitely divisible, where we say that a random variable is *absolutely continuous* if its cumulative distribution function is a continuous function.

1.3 The compound process approximation

In this section we are going to briefly discuss why and how can a Lévy process be approximated by compound Poisson processes. In order to do that, we will do some observations on the proof of theorem 1.11 (Lévy-Khintchine representation theorem) done in [16] and finally introduce how to do the approximation of the Lévy jumps as it is done by Schoutens in [17].

We base the proof on infinitely divisible distribution, as we have already seen the correspondence with Lévy processes. Then, if we are given an infinitely divisible probability measure μ, we choose $t_n \downarrow 0$ and define μ_n by

$$\hat{\mu}_n(z) = \exp\left[{t_n}^{-1}(\hat{\mu}(z)^{t_n} - 1)\right] = \exp\left[{t_n}^{-1}\int_{\mathbb{R}^d\setminus\{0\}}(e^{i\langle z,x\rangle} - 1)\mu^{t_n}(dx)\right]$$

We have that the distribution μ_n is compound Poisson, and by taking $n \to \infty$ one could conclude that $\hat{\mu}(z)$ has the Lévy-Khintchine representation. Is in this part that one firstly finds evidence of what in the literature is called the well-known compound Poisson approximation. Moreover, Sato goes one step beyond and formalizes this result:

Theorem 1.13. *Every infinitely divisible distribution is the limit of a sequence of compound Poisson distributions.*

Now we start with the general procedure of the approximation. Let X be a Lévy process with Lévy triplet $[\gamma, \sigma^2, \nu(dx)]$. First, we discretize the Lévy measure: we choose a small $0 < \epsilon < 1$ and then make a partition of $\mathbb{R} \setminus [-\epsilon, \epsilon]$ of the following form

$$a_0 < a_1 < \ldots < a_k = -\epsilon, \;\; \epsilon = a_{k+1} < a_{k+2} < \ldots < a_{d+1},$$

where $a_i \in \mathbb{R}$. Jumps larger than ϵ are approximated by a sum of independent Poisson processes: we take an independent Poisson process $N^{(i)} = \left\{N_t{}^{(i)}, t \geq 0\right\}$ for each interval, $[a_{i-1}, a_i), 1 \leq i \leq k$ and $[a_i, a_{i+1}), k+1 \leq i \leq d$, with intensity λ_i given by the Lévy measure of the interval. We also need to choose a point c_i, that would be the jump size, in each interval such that the variance of the Poisson process matches the part of the variance of the Lévy process corresponding to this interval.

But now, what if we find jumps smaller than ϵ? We would be looking at the very small jumps, and a first approximation would be by their expected value. Thus, we approximate our Lévy process $X = \{X_t, t \geq 0\}$

by $X^{(d)} = \left\{ {X_t}^{(d)}, t \geq 0 \right\}$, which comprises a Brownian motion W and d independent Poisson processes $N^{(i)}, i = 1, \dots, d$, with intensity parameter λ_i, that as we stated before, depends on the Lévy measure on each interval:

$$
\begin{aligned}
{X_t}^{(d)} &= \gamma t + \sigma W_t + \sum_{i=1}^{d} c_i (N_t^{(i)} - \lambda_i t 1_{|c_i|<1}) \\
\lambda_i &= \begin{cases} \nu([a_{i-1}, a_i)) \text{ for } 1 \leq i \leq k, \\ \nu([a_i, a_{i+1})) \text{ for } k+1 \leq i \leq d, \end{cases} \\
{c_i}^2 \lambda_i &= \begin{cases} \int_{a_{i-1}}^{a_i -} x^2 \nu(dx) \text{ for } 1 \leq i \leq k, \\ \int_{a_i}^{a_{i+1} -} x^2 \nu(dx) \text{ for } k+1 \leq i \leq d, \end{cases}
\end{aligned}
$$

Note that this can be seen as a discretization of the Lévy-Khintchine formula in theorem 1.11 and that these small jumps approximation could be improved by using Brownian Motions.

1.4 Particular processes simulation

In order to do the parameter estimation we need to *a priori* work on estimating the processes so we have the vector of observations $y = (y_1, y_2, \dots, y_n)$. This section will be based on the work by Schoutens in [17]. In general we have at our disposal two different approaches for doing the simulation, as we can either use random number generators or program our own for a certain parameters, and then use properties of the distributions to reach the rest. For each case we will use a time grid of $n = 2^{12}$ intervals. These are all examples with an infinitely divisible characteristic function.

The Gamma Process

A random variable X has a gamma distribution μ with parameters $a > 0$ and $b > 0$: $\Gamma(a, b)$, if its characteristic function is given by:

$$
\phi_\mu(\vartheta) = \left(\frac{b}{b - i\vartheta} \right)^a.
$$

This distribution has the following density function:

$$
f_X(x) = \frac{b^a x^{a-1}}{\Gamma(a)} \exp(-bx), \ \forall x > 0.
$$

We only need a good generator for $\Gamma(a, 1)$ random numbers, as if we have $X \sim \Gamma(a, b)$, then it holds that for $c > 0$, $X/c \sim \Gamma(a, bc)$. For this process, we can easily simulate the path of $G = \{G_t, t \geq 0\}$, where G_t follows a $\Gamma(at, b)$ law, at time points $\{n\Delta t, n = 0, 1, ...\}$. First we need to generate independent $\Gamma(a\Delta t, b)$ random numbers $\{g_n, n \geq 1\}$ and apply:

$$
G_0 = 0, \quad G_{n\Delta t} = G_{(n-1)\Delta t} + g_n, \quad n \geq 1.
$$

To generate the values of g_i we can make use of the *Random Number Generator* or *rng* of software like R, or in the case where $a \leq 1$ (which will likely happen as we assume Δt to be very small), we can follow the **Johnk's Gamma Generator**:

1. Generate two independent uniform random numbers u_1 and u_2.
2. Set $x = {u_1}^{1/a}$ and $y = {u_2}^{1/(1-a)}$.
3. If $x + y \leq 1$ goto step 4, else goto step 1.
4. Generate an Exp(1) random variable, i.e. $\zeta = -\log(u)$, where u is a uniform random number.
5. Return the number $\zeta x/(x+y)$ as the *Gamma*$(a, 1)$ random number.

Again, we test the accuracy of the generator using its theoretical mean and variance, which for a *Gamma*(α, β), $\alpha = $ *shape*, $\beta = $ *rate* we know them to be:

$$\textit{Mean} : \mathbb{E}[X] = \frac{\alpha}{\beta}, \quad \textit{Variance} : Var[X] = \frac{\alpha}{\beta^2}$$

For values of $a = 0.5$, recall that this estimator only works for unitary rate, we obtain the following *RMSE* when testing the mean and variance of 10000 vectors of $n = 2^{12}$ gamma distributed elements each:

$$\textit{Mean} : \textit{RMSE} = 0.011, \quad \textit{Variance} : \textit{RMSE} = 0.029$$

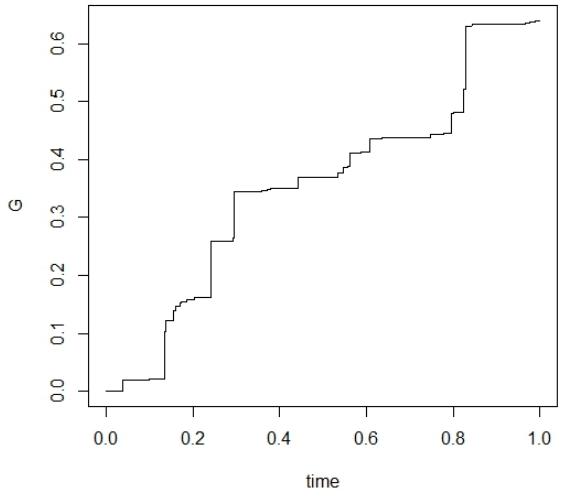

Figure 3: Simulation of a Gamma process with parameters $a = 10$ and $b = 20$.

The TS Process

A non-negative random variable has a tempered stable distribution μ with parameters $a > 0, b \geq 0$ and $\kappa \in]0, 1[$ if its characteristic function is given by:

$$\phi_\mu(\vartheta) = \exp(ab - a(b^{1/\kappa} - 2i\vartheta)^\kappa).$$

This distribution is defined by exponentially tilting a positive and symmetric κ-stable distribution with characteristic function $\phi(\vartheta) = \exp(-a2^{\kappa}\vartheta^{\kappa})$ and density g. More precisely, the density function of this distribution is given by:

$$f(x) = \exp\left(ab - \frac{1}{2}b^{\frac{1}{\kappa}}x\right) g(x).$$

It usually happens with this kind of distributions that the density is only known in the form of series representation. However, there is an important exception when $\kappa = \frac{1}{2}$, as we get an IG distribution, whose density is given by:

$$f(x) = \frac{a\exp(ab)}{\sqrt{2\pi x^3}} \exp\left(-\frac{1}{2}\left(\frac{a^2}{x} + b^2 x\right)\right), \quad \forall x > 0.$$

Furthermore, there are no specific random number generators, so once again we rely on the technique presented by Rosinski [14],[15] called the **path rejection method**. Using this we approximate the path of a TS process $X = \{X_t, 0 \leq t \leq T\}$ with parameters $a > 0,\ b \geq 0$ and $0 < \kappa < 1$ by

$$X(t)^{(M)} = 2\sum_{j=1}^{M} min\left\{\left(\frac{aT}{b_j\Gamma(1-\kappa)}\right)^{\frac{1}{\kappa}}, \frac{e_j v_j^{\frac{1}{\kappa}}}{b^{\frac{1}{\kappa}}}\right\} \mathbf{1}_{Tu_j \leq t}, 0 \leq t \leq T,$$

where $\{e_n, n = 1, 2, ...\}$ is a sequence of independent Exp(1) random numbers, $\{u_n, n = 1, 2, ...\}$, $\{v_n, n = 1, 2, ...\}$, are sequences of independent Uniform(0,1) random numbers and $b_1 < b_2 < ... < b_i < ...$ are the arrival times of a Poisson process with intensity parameter 1. Then, for values of M large enough, we have $X(t)^{(M)} \to X$. Typically, values of M around 10000 are said to give very reasonable approximations.

However, this is a simulation process of order $\mathcal{O}(nM)$, where again, n is the number of time intervals and M the convergence parameter, meaning that this is not a fast algorithm to follow.

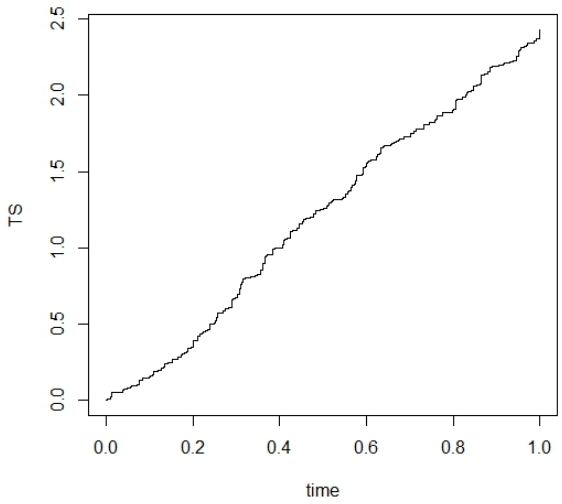

Figure 4: Simulation of a TS process with parameters $a = 10$ and $b = 2$ and $\kappa = 0.15$.

The IG Process

Let $T^{(a,b)}$ be the first time a standard Brownian Motion with drift $b > 0$ reaches the positive level $a > 0$, this is a random time that follows an $IG(a, b)$. It has characteristic function

$$\phi(\vartheta) = \exp(-a(\sqrt{-2iu + b^2} - b)),$$

and density function

$$f(x) = \frac{a}{\sqrt{2\pi}} \exp(ab) x^{-3/2} \exp(-\frac{1}{2}(a^2 x^{-1} + b^2 x)),\ x > 0.$$

To simulate a sample path of an Inverse Gaussian process $I = \{I_t, t \geq 0\}$, $I_t \sim IG(at, b)$, we discretize the time points in the grid $\{n\Delta t, n = 0, 1, ...\}$ and generate independent $IG(a\Delta t, b)$ random numbers $\{i_n, n \geq 1\}$ to follow this scheme

$$I_0 = 0,\ \ I_{n\Delta t} = I_{(n-a)\Delta t} + i_n,\ n \geq 1.$$

Regarding the simulation of the IG numbers we have two approaches. As we have just simulated a TS process, we could use one of its properties already mentioned: set $\kappa = 0.5$ and follow again the path rejection method. However, not only the simulation time will be slow, but even when setting the parameter M to 10000 as said by Schoutens, the results are not really precise. Given that the theoretical mean and variance are

$$\textit{Mean} : \mathbb{E}[X] = \frac{a}{b},\ \ \textit{Variance} : Var[X] = \frac{a}{b^3},$$

For $a = 1$, $b \simeq 49$ and 25 samples, we obtain:

$$\textit{Mean} : \textit{RMSE} = 1.367559,\ \ \textit{Variance} : \textit{RMSE} = 0.6424164,$$

And has taken 1h to do the calculations with an *IntelCore i7* and CPU of 2.60GHz. The best alternative would be using an IG *rng* proposed by Michael *et al.* (1976) [11], that consists of the following steps in order to get an $IG(a, b)$:

1. Generate a standard Normal random number v.
2. Set $y = v^2$
3. Set $x = (a/b) + y/(2b^2) - \sqrt{4aby + y^2}/(2b^2)$.
4. Generate a uniform random number u.
5. If $u \leq (a + xb)$, then return x as the $IG(a, b)$. Otherwise, return $a^2/(b^2 x)$.

Using this method, we obtain the following estimation results:

$$\textit{Mean} : \textit{RMSE} = 0.0011,\ \ \textit{Variance} : \textit{RMSE} = 1.1696e - 09,$$

which yields a better and considerable faster generation.

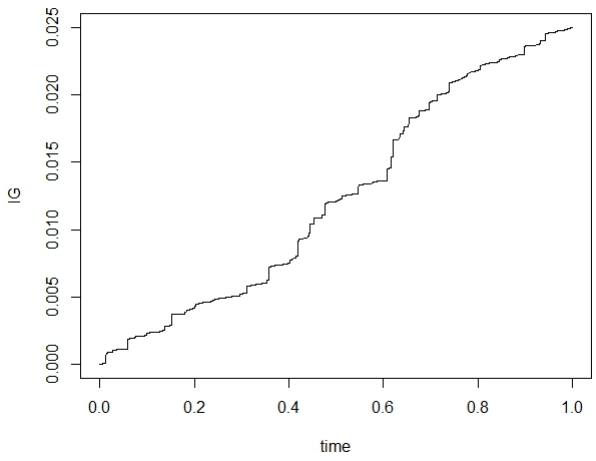

Figure 5: Simulation of an IG sample path with $a = 1, b \simeq 49$

The NIG Process

A random variable has a Normal Inverse Gaussian (NIG) distribution with parameters $\alpha > 0$, $-\alpha < \beta < \alpha$ and $\delta > 0$, has a characteristic function

$$\phi(\vartheta) = \exp(-\delta(\sqrt{\alpha^2 - (\beta + iu)^2} - \sqrt{\alpha^2 - \beta^2})),$$

and density function

$$f(x) = \frac{\alpha\delta}{\pi} \exp(\delta\sqrt{\alpha^2 - \beta^2} + \beta x) \frac{K_1(\alpha\sqrt{\delta^2 + x^2})}{\sqrt{\delta^2 + x^2}},$$

where $K_\lambda(x)$ denotes the modified Bessel function of the third kind with index λ. A NIG Process $X^{(NIG)} = \{X_t{}^{(NIG)}, t \geq 0\}$ can be simulated as a time-changed Brownian Motion where the drift is done by an IG process $I = \{I_t, t \geq 0\}$ with parameters $a = 1$ and $b = \delta\sqrt{\alpha^2 - \beta^2}$. We have

$$X^{(NIG)} = \beta\delta^2 I_t + \delta\, W_{I_t}. \tag{5}$$

Thus, a sample path of the NIG process can be obtained by sampling a standard Brownian Motion and an IG process. Note that the Wiener process is not sampled on t, but on I_t. This means that when computing its values, we first need to generate standard Wiener process intervals $dW_t \sim \sqrt{dt}\mathcal{N}(0, 1)$, then times $\sqrt{I_t}$. Another important fact is the weight of the drift given by β when δ is small compared to it. In this case, as we have a negative drift, the Wiener process only gives a small volatility to a mainly decreasing process, as I_t is positive. For the simulation, we followed this scheme:

1. Generate an IG sample path I_t.
2. Generate a standard Brownian Motion in the same grid W_t.
3. Calculate $W_{I_t} = W_t * \sqrt{I_t}$: Brownian Motion in the I_t steps.
4. Apply (5).

Comparing the theoretical moments against the sample ones to test the estimator we obtain:

$$\textit{Mean} : \mathbb{E}[X] = \delta\beta/\sqrt{\alpha^2 - \beta^2}, \quad \textit{Variance} : Var[X] = \alpha^2\delta(\alpha^2 - \beta^2)^{-3/2},$$
$$\textit{Mean} : \textit{RMSE} = 2.525e-05, \quad \textit{Variance} : \textit{RMSE} = 2.544e-06$$

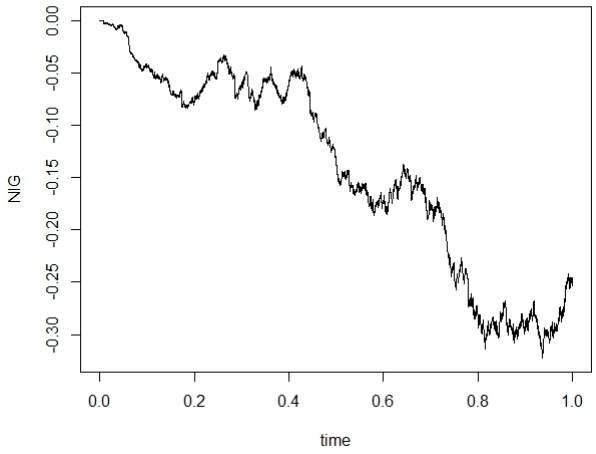

Figure 6: Simulation of a NIG sample path with $\alpha = 50,\ \beta = -10,\ \delta = 1$

2. Parameter Estimation

Our prime objective is the parameter estimation of stochastic processes and furthermore the exploration of a method that shines in the cases where we can't directly work with the density function of the processes distributions. However, we will want to compare it with a more wide known method for parameter estimation: the **Maximum Likelihood Estimation** (MLE).

It is based on using n *i.i.d.* observations ruled by a probability density function f of a certain family of distributions $\{f(\cdot|\theta), \theta \in \Theta\}$, from which the vector of parameters θ is unknown. Then, one will work with the join density function

$$f(x_1, x_2, \dots, x_n \mid \theta) = f(x_1 \mid \theta) \times f(x_2 \mid \theta) \times \cdots \times f(x_n \mid \theta),$$

in order to compute the so-called likelihood function

$$\mathcal{L}(\theta\,;\, x_1, \dots, x_n) = f(x_1, x_2, \dots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta).$$

where the observations will work as fixed parameters that are used to obtain the vector θ through the maximization of $\mathcal{L}$ or its logarithm, which usually eases the process computationally, so that $\hat{\theta} = \arg\max l_n(\theta)$.

To test the computational power of this algorithm, we will work with the R software library *fBasics*, which has a built-in implementation of the MLE for the NIG distribution. We are going to focus on its accuracy and time lapse for different number of observations. The real parameters we are using for the simulation are: $\alpha = 2,\ \beta = 1,\ \delta = 1,\ \mu = 0$, and the results are available in Table 1. The estimation error is calculated using:

$$\textit{error} = \sqrt{\sum_i (\theta_i - \hat{\theta}_i)^2}$$

n° of obs.	Estimations	Error	Time
$n = 500$	$\alpha = 2.4127,\ \beta = 1.3804,\ \delta = 1.0042,\ \mu = -0.1158$	0.573	0.19 s
$n = 1000$	$\alpha = 1.8047,\ \beta = 0.8253,\ \delta = 0.9546,\ \mu = 0.05314$	0.271	0.231 s
$n = 10000$	$\alpha = 1.9637,\ \beta = 0.9597,\ \delta = 1.0230,\ \mu = 0.00657$	0.059	1.017 s
$n = 100000$	$\alpha = 2.0186,\ \beta = 1.0078,\ \delta = 1.0042,\ \mu = -0.0064$	0.0216	11.434 s

Table 1: MLE for a NIG process.

We can see that we need at least 10000 observations so that the estimates start to resemble the real values. Also, note that this method would not work with processes like the Tempered Stable, as we can only know the values of the density through numerical approximations, but we do have a closed form for its characteristic function.

3. The Ornstein-Uhlenbeck process

Definition 3.1. A stochastic process is said to be an Ornstein-Uhlenbeck process if it satisfies the following differential equation:

$$\begin{cases} dX(t) = -\lambda X(t)dt + \sigma dW(t) \\ X(0) = X_0 \end{cases} \tag{6}$$

where λ is a strictly positive intensity parameter and X_0 is an independent random variable. As we said, a Brownian Motion is just a base case of a bigger family of stochastic processes, and thus we can generalize this process by allowing W to be a Lévy process $\mathbf{Z}$, that will be referred as the background driving Lévy process (BDLP). This generalization allows us to design a self-decomposable law D and then find a BDLP Z that matches its distribution. Thus, we will call this processes *D-OU* processes.

The Ornstein-Uhlenbeck process can be seen as the description of the velocity of a Brownian particle under the influence of friction, and it is one of the models used in finance for calculating interest rates or currency exchange rates and were introduced by Barndorff-Nielsen and Shephard [4] as a model to describe volatility. It is the only process that satisfies these three conditions at the same time:

1. **Stationary process**: the joint probability distribution does not change when shifted in time:

   $$F_X(x_{t_1+\tau}, \dots, x_{t_k+\tau}) = F_X(x_{t_1}, \dots, x_{t_k})$$

   where F_X denotes the cumulative distribution function of the joint distribution of $\{X_t\}$.

2. **Gaussian process**: for every finite set of indices $t_1, \dots, t_k$, $X_{t_1,\dots,t_k} = (X_{t_1}, \dots, X_{t_k})$ is a multivariate Gaussian random variable.

3. **Markov property**: memoryless, meaning that if we know the filtration of the process until time s, we are only interested on what happened at time s, but not before. More precisely:

   $$\mathbb{P}(X_t \in A|\mathcal{F}_t) = \mathbb{P}(X_t \in A|X_s)$$

As shown in [9], given the linear stochastic differential equation:

$$\begin{cases} dX_t = [A(t)X_t + a(t)]dt + \sigma(t)dW_t,\ 0 \le t < \infty \\ X(0) = X_0 \end{cases} \tag{7}$$

there is a unique strong solution given by:

$$dX_t := \Phi(t)\left[X_0 + \int_0^t \Phi^{-1}(s)a(s)ds + \int_0^t \Phi^{-1}(s)\sigma(s)dW_s\right],\ 0 \le t < \infty \tag{8}$$

where the matrix function $\Phi(t)$ is the fundamental solution of the homogeneous equation. This solution can be easily verified by using the Itô's Lemma, the analogous version of the chain rule in stochastic calculus.

Lemma 3.2. *(Itô's Lemma) If $X(t)$ is an Itô process: a continuous in time stochastic process satisfying:*

$$dX_t = \mu_t\, dt + \sigma_t\, dW_t$$

and $f = f(X, t)$ is a differentiable function of X_t and t, then:

$$df = \left(\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial x} + \frac{\sigma_t^2}{2} \frac{\partial^2 f}{\partial x^2} \right) dt + \sigma_t \frac{\partial f}{\partial x} \, dW_t \tag{9}$$

Following this approach one can find that the solution of (6) is

$$X_t = exp(-\lambda t) \left(X_0 + \sigma \int_0^t exp(\lambda s) \, dW(s) \right)$$

where we used $\Phi(t) = \exp(-\lambda t)$, $A(t) = -\lambda$ and $a \equiv 0$. However, another special characteristic of this model is that the same result will be obtained by just using the usual calculus techniques for solving differential equations. For the calculation of the expectancy, variance and covariance we will refer again to results showed in [9]:

- $m(t) = \Phi(t) \left[m(0) + \int_0^t \Phi^{-1}(s) a(s) ds \right]$.
- $\rho(s, t) = \Phi(s) \left[Var(0) + \int_0^{\min\{s,t\}} \Phi^{-1}(u) \sigma(u) (\Phi^{-1}(u) \sigma(u))^T du \right]$.

And we obtain:

- $\mathbb{E}[X(t)] = m(t) = \exp(-\lambda t) \mathbb{E}[X_0]$.
- $Var(X(t)) = \rho(t, t) = \frac{\sigma^2}{2\lambda} + \left(Var(X_0) - \frac{\sigma^2}{2\lambda} \right) \exp(-2\lambda)$.
- $Cov(X(s), X(s+t)) = \rho(s, s+t) = \left(Var(X_0) + \frac{\sigma^2}{2\lambda} (\exp(2\lambda s) - 1) \right) \exp(-\lambda(2s + t))$.

3.1 Relation with ARMA models

In this section we are going to briefly introduce the work presented on [1], in which they show the link between the time continuous Ornstein-Uhlenbeck process to analogue time discrete ARMA processes.

Definition 3.3. (*The class of ARMA processes*) Consider a weak white noise, $W_t \sim W\mathbf{N}(0, \sigma^2)$, and let integers $p \geq 1$ and $q \geq 1$. A time series X_t is

- **AR**(p) (autoregressive of order p) if

$$X_t = W_t + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p}.$$

- **MA**(q) (moving average of order q) it

$$X_t = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + ... + \theta_q W_{t-q}.$$

- **ARMA**(p, q) (autoregressive and moving average of order p, q)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + ... + \theta_q W_{t-q}.$$

where $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ are all real numbers.

Then if we consider the OU process X sampled at equidistant times $h\tau : h = 0, 1, 2, \dots, n, \tau > 0$, the series $X_h = x(h\tau)$ obeys an AR(1): $X_h = e^{-\lambda\tau} + \mathbf{A}_{h+1}$, where $\mathbf{A}_{h+1} = \int_{h\tau}^{(h+1)\tau} e^{-\lambda((h+1)\tau - s)}\, dZ(s)$ and a Lévy process Z, that constitutes the stochastic innovation. Thus, we can consider the OU process as a continous-time interpolation of an AR(1) process. One could also iterate the *OU* operator to get a generalized OU(p) process which relates in discrete time to a ARMA($p, p-1$) process as shown in [1].

3.2 Processes of Lévy-OU type

In this section we are going to present the results that will let us construct the D-OU processes, following the work of Valdivieso *et al.* in [20]. Let $Z = \{Z_t\}$ be a univariate Lévy process, it can be decomposed as follows: $Z = X_1 + X_2 + X_3$, where X_1 is a Wiener process with drift 0 and volatility σ_0, X_2 is a compound Poisson process (stochastic process with random jumps of random size) and X_3 is a square integrable pure jump martingale that a.s has a countable number of jumps on a finite time interval. For this reason, we say that a Lévy process Z has a generating triplet $(\sigma_0, \gamma_0, \nu_0)$, where ν_0 is the Lévy measure of the process. A stochastic process $X = X(\mathrm{t})$ is said to be a process of Ornstein-Uhlenbeck type generated by $(\sigma_0, \gamma_0, \nu_0, \lambda)$ if it is càdlàg and satisfies the stochastic differential equation

$$\begin{cases} dX(t) = -\lambda X(t)\, dt + \sigma\, dZ(\lambda t) \\ X(0) = X_0 \end{cases}, \tag{10}$$

where X_0 is a random variable independent of the BDLP Z. Then, as shown in [12] one has to use the theory of integration of left-continuous predictable processes with respect to *semimartingales*, and particularly to Lévy processes, we obtain the following integral form of the stochastic differential equation:

$$X(t) = \exp(-\lambda t)\left(X_0 + \int_0^t \exp(\lambda s)\, dZ(\lambda s)\right), \tag{11}$$

for any $t \geq 0$. Also, for $\Delta \geq 0$ we can define recursively

$$X(t+\Delta) = \exp(-\lambda \Delta)\left(X_t + \int_t^{t+\Delta} \exp(\lambda s)\, dZ(\lambda s)\right), \tag{12}$$

from where we can observe that X is a Markov process, as the future values of X only depend on the actual value of the process, not taking in account the values before the time t.

Definition 3.4. A *semimartingale* is a process that can be decomposed as the sum of a *local martingale* and an adapted *finite variation* process, which are processes whose paths are right continuous and have a finite total variation over every compact time interval, with probability one.

The basic definition of a Lévy driven OU process considers that Z_t is a subordinator: a Lévy process with no Brownian part, nonnegative drift and only positive increments. As $X_0 > 0$, it means that the process is strictly positive and bounded from below by the deterministic function $X_0 \exp(-\lambda t)$. However, the drift term can easily be included in the SDE

$$dX(t) = (\alpha - \lambda)X(t)\, dt + \sigma\, dZ(\lambda t)$$

as we can transform this equation in the form of (10) by defining a new BDLP $\hat{Z}_t = Z_t + \lambda^{-1}\alpha t$. If we define the Lévy functional as $Z^*(\Delta) = \int_0^{\lambda\Delta} \exp(s)dZ(s)$, we have:

Proposition 3.5. *For any $t \geq 0$ and $\Delta \geq 0$*

$$Z^*(\Delta) \overset{d}{=} \int_0^{\Delta} \exp(\lambda s)dZ(\lambda s) \overset{d}{=} \exp(-\lambda t)\int_t^{t+\Delta} \exp(\lambda s)dZ(\lambda s),$$

where $\overset{d}{=}$ means equality in distribution.

Proof [20]: The first equality holds for simple predictable integrand processes and then, by density arguments of the exponential function. We have $Z^*(\Delta/\lambda) = \int_0^{\Delta} \exp(s)dZ(s)$, and by linearity of the functional $\frac{1}{\lambda}Z^*(\Delta) = Z^*(\Delta/\lambda)$. We also have, for $c \in \mathbb{R}$, that $cZ^*(\Delta) = c\int_0^{\lambda\Delta} \exp(s)dZ(s) = \int_0^{\lambda\Delta} \exp(s)dZ(cs)$ by linearity of the differential operator. Finally, using that $\exp(s\lambda) = \exp(s)\exp(\lambda)$, we multiply by λ in both sides of the equation $\frac{1}{\lambda}Z^*(\Delta) = \int_0^{\Delta} \exp(s/\lambda)dZ(s/\lambda)$, and as $\exp(\lambda)$ is nothing but a constant we multiply by it on the rhs and get: $Z^*(\Delta) \overset{d}{=} \int_0^{\Delta} \exp(\lambda s)dZ(\lambda s)$, as making times a constant only is a translation of the density function.

For the second equality we need first to introduce an important result shown in [12], the integration by parts for semimartingales.

Theorem 3.6. *(Integration by parts) Let X and Y be two semimartingales, then for all $t \geq 0$, XY is a semimartingale and:*

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + [X, Y]_t,$$

where $[X, Y]$ is the quadratic covariation of X and Y, which is a bilinear and symmetric operator defined by:

$$[X, Y]_t = \lim \sum_i (X(t_{i+1}) - X(t_i)(Y(t_{i+1}) - Y(t_i)).$$

However, for the proof we will need to write this theorem in another time interval: $[t, t+\Delta]$, so we end up with:

$$X_{t+\Delta} Y_{t+\Delta} = X_t Y_t + \int_t^{t+\Delta} X_s dY_s + \int_t^{t+\Delta} Y_s dX_s + [X, Y]_{t+\Delta}, \tag{13}$$

which is consistent if we consider X and Y to be Markov processes. Now, using this expression we can express $\int_t^{t+\Delta} \exp(\lambda s)dZ(\lambda s)$ as

$$e^{\lambda(t+\Delta)}Z(\lambda(t+\Delta)) - e^{\lambda t}Z(\lambda t) - e^{\lambda t}\int_{\lambda t}^{\lambda(t+\Delta)} e^s\, Z(s)ds.$$

For this we use (13), where X and Y are respectively the exponential density and the Lévy process Z (both of them holding the Markov property). Note that in this case the covariation is equal to zero. In particular:

Theorem 3.7. *Given X, Y two stochastic processes. If either X or Y is a finite variation process and either X or Y is continuous, then $[X, Y]_t = 0$.*

We have both X and Y being finite variation processes and the exponential, continuous. The proof for this and further information about stochastic properties and definitions we are using can be found in [10]. At

this point, the association of the elements in the expression and the ones in (13) is mostly self-explanatory. We just need to work a bit more on the last one:

First, we apply a translation to the integral,use again the linearity of the differential operator and solve an integral only involving the exponential so that $\int_{\lambda t}^{\lambda(t+\Delta)} e^s\ Z(s)ds$ equals to

$$e^{\lambda t}\int_0^{\lambda\Delta} e^s\ Z(s+\lambda t) \overset{d}{=} e^{\lambda t}\left[\int_0^{\lambda\Delta} e^s\ Z(s)ds + Z(\lambda t)(e^{\lambda\Delta}-1)\right],$$

And so

$$\int_t^{t+\Delta} e^{\lambda s}dZ(\lambda s) \overset{d}{=} e^{\lambda(t+\Delta)}Z(\lambda\Delta) + e^{\lambda(t+\Delta)}Z(\lambda(\lambda t)) - e^{\lambda t}\left[\int_0^{\lambda\Delta} e^s\ Z(s)ds + Z(\lambda t)(e^{\lambda\Delta}-1)\right] =$$

$$= e^{\lambda(t+\Delta)}Z(\lambda\Delta) - e^{\lambda t}\int_0^{\lambda\Delta} e^s\ Z(s)ds.$$

If we now apply a change of variable in the integral we are left with: $s \to \lambda s$, we have in the upper limit of integration $\phi(\Delta) = \lambda\Delta$ so $\phi'(\Delta) = \lambda$:

$$\int_0^{\lambda\Delta} e^s\ Z(s)ds = \int_0^{\Delta} s^{\lambda s}Z(\lambda s)\lambda ds,$$

and doing the usual integration by parts, where $v' = \lambda e^{\lambda s}$ and $u = Z(\lambda s)$ we get

$$e^{\lambda\Delta}Z(\lambda\Delta) - \int_0^{\Delta} e^{\lambda s}dZ(\lambda s).$$

Finally we can conclude with the equality we were looking for:

$$e^{-\lambda t}\int_t^{t+\Delta} e^{\lambda s}dZ(\lambda s) \overset{d}{=} \int_0^{\Delta} e^{\lambda s}dZ(\lambda s). \quad \square$$

We will understand any distribution μ as a probability distribution on $\mathbb{R}$. The characteristic and cumulant function of μ will be denoted, respectively, by ϕ_μ and $\mathcal{C}_\mu = \log(\phi_\mu)$.

Regarding to self-decomposability, Sato has also established in [16] a fundamental result about the stationarity of a process of Ornstein-Uhleneck type.

Proposition 3.8. *If X is a process of Ornstein-Uhlenbeck type generated by $(\sigma_0, \gamma_0, \nu_0, \lambda)$ such that*

$$\int_{|x|>2} \log(|x|)d\nu_0(x) < \infty, \tag{14}$$

then X has a unique self-decomposable stationary distribution μ.

Conversely, for any $\mu > 0$ and any self-decomposable distribution D, there exists a unique triplet $(\sigma_0, \gamma_0, \nu_0)$ satisfying (14) and a process of OU type X generated by $(\sigma_0, \gamma_0, \nu_0, \lambda)$ such that D is the stationary distribution of X.

Thanks to this Proposition, given three different self-decomposable distributions D_1, D_2, D_3,we will be able to create the D_i-OU type processes, $i \in \{1, 2, 3\}$, from which we are going to do the simulation. From (12), the autocorrelation function of these processes takes the form

$$\rho(\Delta) = \frac{Cov(X(t), X(t+\Delta))}{\sqrt{Var(X(t))\,Var(X(t+\Delta))}} = \exp(-\lambda\,|\Delta|), \tag{15}$$

for any $t \geq 0$ and $\Delta \in \mathbb{R}$. If (σ, γ, ν) denotes the Lévy triplet of the stationary distribution D, then $(\sigma_0, \gamma_0, \nu_0)$ satisfy

$$\sigma_0 = 2\sigma \quad \textit{and} \quad \gamma_0 = \gamma - \int_{-\infty}^{-1} \omega(x)\,dx - \int_1^{\infty} \omega(x)\,dx,$$

where ω is the Lévy density associated to ν_0. In [3] it is shown that if the Lévy density associated to ν, u, is differentiable, then ω satisfies

$$\omega(x) = -u(x) - xu'(x).$$

We need to introduce one last result (proved also in [3]) which is a key relation among the cumulants of D, Z and the Lévy functional $Z^*(\Delta)$.

Proposition 3.9. *For any $\Delta > 0$ and $\vartheta \in \mathbb{R}$:*

$$\mathcal{C}_{Z^*(\Delta)}(\vartheta) = \log(\mathbb{E}[\exp(i\vartheta Z^*(\Delta))]) = \lambda \int_0^{\Delta} \mathcal{C}_{Z(1)}(\vartheta \exp(\lambda s))\,ds,$$

where $\mathcal{C}_{Z(1)}(\vartheta) = \vartheta \frac{d\mathcal{C}_D(\vartheta)}{d\vartheta}$.

We now introduce some important D-OU processes.

3.3 The gamma $\Gamma(a, b)$-OU process

A random variable X has a gamma distribution μ with parameters $a > 0$ and $b > 0$: $\Gamma(a, b)$, if it has the following characteristic function:

$$\phi_\mu(\vartheta) = \left(\frac{b}{b - i\vartheta}\right)^a.$$

In [6] it is proved that this is a self-decomposable distribution. Hence, by Proposition 3.5 the $\Gamma(a, b)$-OU process is well defined. Moreover, the BDLP Z of this process has, using Proposition 3.9, the following cumulant distribution function:

$$\mathcal{C}_{Z(1)}(\vartheta) = \frac{a\vartheta i}{b - i\vartheta} = a\left(\frac{b}{b - i\vartheta} - 1\right).$$

Recall that we defined the cumulant function as $\mathcal{C}_\mu = \log(\phi_\mu)$, so using this and calculating the derivative by ϑ we get to this result. Consequently, $Z = \{Z(t)\}$ is a compound Poisson process with intensity parameter a and associated exponential distribution with parameter b. Heuristically this can be shown just by thinking about a couple of aspects: first of all, the characteristic function comes from an expectancy, and second, we want for Poisson processes to hold $N_0 = 0$. Then, the above cumulant function is another characteristic function: i.e. can be thought as an expected value which would be a times the exponential characteristic function minus 1 (this way at time 0 we have value 0), so we have a times the expected value of an exponential distribution, where a is the expected number of jumps as we said that the compound Poisson had this intensity parameter. However, we will later show that not only this BDLP can be estimated through a compound Poisson, but for all Lévy process this property holds.

3.4 The tempered stable $TS(\kappa, a, b)$-OU process

A non-negative random variable has a tempered stable distribution μ with parameters $a > 0, b \geq 0$ and $\kappa \in]0, 1[$ if its characteristic function is given by:

$$\phi_\mu(\vartheta) = \exp(ab - a(b^{1/\kappa} - 2i\vartheta)^\kappa).$$

As we have already discussed, μ is infinitely divisible, and so we can define a $TS(\kappa, a, b)$-Lévy process, which has Lévy density

$$u(x) = \frac{a\kappa 2^\kappa x^{-(\kappa+1)}}{\Gamma(1-\kappa)} \exp\left(-\frac{b^{\frac{1}{\kappa}} x}{2}\right).$$

In Barndorff-Nielsen and Shepard (2002) [5] it is shown that the BDLP $Z = \{Z(t)\}$ of a $TS(\kappa, a, b)$-OU process admits for any $t \geq 0$ the decomposition:

$$Z(t) \stackrel{d}{=} I(t) + \sum_{j=1}^{N(t)} Y_j,$$

where $\mathbf{I} = \{I(t)\}$ is a $TS(\kappa, \kappa a, b)$-Lévy process, $\{N(t)\}$ is a Poisson process with intensity parameter $ab\kappa$ and $Y_1, Y_2, \ldots$ are independent and identically distributed $\Gamma(1-\kappa, \frac{b^{\frac{1}{\kappa}}}{2})$ random variables.

3.5 The normal inverse Gaussian $NIG(a, \beta, \delta)$-OU process

A random variable X has a normal inverse Gaussian distribution μ with parameters α, β and δ if its characteristic function is given by:

$$\phi_\mu(\vartheta) = \exp(-\delta(\sqrt{\alpha^2 - (\beta + iu)^2} - \sqrt{\alpha^2 - \beta^2})),$$

showing that μ is infinitely divisible and closed under convolutions if α and β are kept fixed. One interesting property of a $NIG(\alpha, \beta, \delta)$-Lévy process $Y = \{Y_t\}$ which we used for its simulation is that it can be expressed as an inverse Gaussian time-changed Brownian motion:

$$Y(t) = \beta\delta^2 IG(t) + \delta\, W(IG(t)).$$

Moreover, Y has no Brownian motion part, drift $\gamma = \frac{2\delta\alpha}{\pi} \int_0^1 \sinh(\beta x)\ K_1(\alpha x) dx$ and Lévy measure $\nu(B) = \int_B u(x)dx = \int_B \frac{\delta\alpha}{\pi|x|} K_1(\alpha\,|x|) \exp(\beta x) dx$. Halgreen has shown in [6] that μ is self-decomposable, so the $NIG(\alpha, \beta, \delta)$-OU process is well defined. This process has a pure-jump BDLP with density

$$w(x) = (1 - \beta x)u(x) + \frac{\delta\alpha^2}{\pi} K_0(\alpha\,|t|) \exp(\beta x).$$

An important difference with the so far analyzed D-OU processes, is that while the other have paths with bounded variation, the NIG-OU process are of unbounded variation with infinitely many up and down jumps in any finite interval.

4. Likelihood inference for a D-OU process

Let $X = X(t)$ be a D-OU process, where the stationary distribution D depends on an unknown parameter $\theta \in \mathbb{R}^m$. Suppose we are interested in estimating θ based on a set of n+1 observations $x_0, x_1, \dots, x_n$ coming from a sample $X(0), X(t_1), X(t_2), \dots, X(t_n)$ of X.

The likelihood function $L(\theta)$ is the joint density of the sample. By the Markov property, the densities of each X_{t_i} conditioned to all past values, only depends on the last one of them, i.e. $f_{X(t_k)|X(t_{k-1})=x_{k-1}}(x_k)$. Thus, the likelihood function of this sample can be written as

$$L(\theta) = f_{X(0)}(x_0) \prod_{k=1}^{n} f_{X(t_k)|X(t_{k-1})=x_{k-1}}(x_k)$$

If we recall now the recursive expression of X_t showed in (12), and by means of the Proposition 3.5, we obtain the following autoregressive scheme:

$$X(t_{k-1}+\Delta_k) = X(t_{k-1}+t_k-t_{k-1}) = X(t_k) = \exp(-\lambda\Delta_k)\left(X(t_{k-1}) + \exp(-\lambda t_{k-1})\int_{t_{k-1}}^{t_k} \exp(\lambda s)\, dZ(\lambda s)\right)$$

$$\stackrel{d}{=} \exp(-\lambda\Delta_k)(X(t_{k-1}) + Z^*(\Delta_k)),$$

where $t_0 = 0$, $\Delta_k = t_k - t_{k-1}$ *and* $Z^*(\Delta_k) = \int_0^{\lambda\Delta_k} \exp(s)\, dZ(s)$.

Now we are able to relate the distribution of $X(t_k)$ with the distribution of the Lévy functional:

$$\mathbb{P}(X(t_k) \le x_k \mid X(t_{k-1}) = x_{k-1}) = \mathbb{P}(Z^*(\Delta_k) \le \exp(\lambda\Delta_k)x_k - x_{k-1}),$$

and if $Z^*(\Delta_k)$ is a continuous random variable

$$f_{X(t_k)|X(t_{k-1})=x_{k-1}}(x_k) = \exp(\lambda\Delta_k) f_{Z^*(\Delta_k)}(\exp(\lambda\Delta_k)x_k - x_{k-1}).$$

We can assume, without loss of generality, a regular grid $\Delta_k = \Delta$ with $k = 1, 2, \dots, n$ and an absolutely continuous distribution for $Z^*(\Delta_k)$. With these assumption, the likelihood functions becomes

$$L(\theta) = f_{X(0)}(x_0) \exp(n\lambda\Delta) \prod_{k=1}^{n} f_{Z^*(\Delta)}(\exp(\lambda\Delta)x_k - x_{k-1}),$$

so in order to perform the maximum likelihood inference we just need to know the densities of D and $Z^*(\Delta)$.

Definition 4.1. Given $f \in L^1(\mathbb{R}^n)$, i.e. f is an integrable function, we denote by $\widehat{f}$ the Fourier transform of f, defined by

$$\widehat{f}(\nu) = \int_{\mathbb{R}^n} \exp(i\langle\nu, x\rangle) f(x),$$

For a real random variable $X \in \mathbb{R}^n$ with distribution μ, its characteristic function is defined by

$$\phi_X(\nu) = \mathbb{E}[\exp((i\langle\nu, x\rangle)] = \int_{\mathbb{R}^n} \exp(i\langle\nu, x\rangle)\mu(dx).$$

Therefore, if μ has density function f: $\mu(dx) = f(x)dx$, then $f \in L^1(\mathbb{R}^n)$ and

$$\phi_X = \widehat{f}$$

Theorem 4.2. *(The Inversion Formula) Suppose that $f, \widehat{f}$ belong to $L^1(\mathbb{R}^n)$. Then the equality*

$$f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \exp(-i\langle \nu, x\rangle)\widehat{f}(\nu)d\nu,$$

holds for almost all $x \in \mathbb{R}^n$.

Thanks to this formula and Proposition 3.9, we can easily calculate the characteristic function of $Z^*(\Delta)$, and thus its density function, via a discrete fast Fourier transform. To simplify matters at first, let us assume for the moment that the intensity parameter λ is known and we are interested in the vector parameter θ indexing the stationary distribution D. Then we propose to consider the following steps:

- Find an initial estimator $\hat{\theta}$ of θ. Using (12), we propose to consider the n independent and identically distributed random variables

$$Y_k = \int_{\lambda(k-1)\Delta}^{\lambda k\Delta} \exp(s)dZ(s) = \exp(\lambda\Delta)\left(X(k\Delta) - X((k-1)\Delta)\right) \tag{16}$$

  with $k = 1, 2, \dots, n$, and obtain $\hat{\theta}_0$ by matching for this sample the empirical and theoretical mean, variance, bias and kurtosis (depending on the number of parameters we need to do the initial estimate). In practice, this initial estimation can be improved by means of the generalized method of moments.

- Use Proposition 3.9 to obtain the cumulant function of the BDLP Z and the characteristic function of the Lévy functional $Z^*(\Delta)$, which can be used to evaluate all of the theoretical moments above.

- Use the preceding characteristic function and the classical or fractional discrete Fast Fourier transform to evaluate the density function $f_{Z^*(\Delta)}$.

- Use a numerical method to optimize the likelihood function. Although a global optimization can be used, like the differential evolution algorithm, it is better for speeding purposes to start with a good estimator $\hat{\theta}_0$ and then try a local search algorithm around it.

If λ is not given, with (15) we can estimate it by

$$\hat{\lambda}_0 = -\frac{\log(\text{acf}(1))}{\Delta},$$

or by solving

$$\hat{\lambda}_0 = \arg\min_{\lambda} \sum_{k=1}^{m} (\text{acf}(k) - \exp(-\lambda k\Delta))^2,$$

where $\text{acf}(k)$ denotes the empirical autocorrelation function of lag k based on the data $x_0, x_1, \dots, x_n$. As a criterion, we would truncate the sum above as soon as the empirical autocorrelation function reaches the

level $\frac{1.96}{\sqrt{n}}$. This is the 95% upper confident limit interval for the autocorrelation function of a white noise process, as it is distributed as a Gaussian, and 1.96 is the 95% quantile.

Once $\hat{\lambda}_0$ is obtained, there are two possible methods. The first it is nothing but a plug in method consisting in following the above scheme after substituting λ with $\hat{\lambda}_0$. The second is a proper likelihood method that considers $(\hat{\theta_0}(\hat{\lambda}_0), \hat{\lambda}_0)$ as an initial estimator in the search of (θ, λ) that maximizes the likelihood function. We will base our results on this last method.

Before continuing, we will make reference to the empirical mean, variance, bias and kurtosis of the random sample:

- $\bar{Y} = \frac{\sum_{k=1}^{n} Y_k}{n}$
- $S^2 = \frac{\sum_{k=1}^{n} (Y_k - \bar{Y})^2}{n}$
- $B = \frac{\sum_{k=1}^{n} (Y_k - \bar{Y})^3}{nS^3}$
- $K = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^4}{nS^4}$

4.1 Inference for a $\Gamma(a, b)$-OU process

As we have already explained, the initial estimators are found by matching the theoretical and empirical mean and variance. We use that the mean is a linear operator and that $X_t \sim \Gamma(a, b), \forall t$, where the mean and variance of the gamma distribution are $\frac{a}{b}$ and $\frac{a}{b^2}$ respectively:

$$\mathbb{E}[Y_1] = \mathbb{E}[e^{\lambda \Delta} X(\Delta) - X_0] = \frac{a(\exp(\lambda \Delta) - 1)}{b}, \text{ and } Var(Y_1) = \frac{a(\exp(2\lambda \Delta) - 1)}{b^2},$$

as $Var(aX) = a^2 Var(X)$. Now we just solve the system to get $\hat{\theta}_0 = (\hat{a}_0, \hat{b}_0)$ and yields to:

$$\hat{a}_0 = \frac{\bar{Y} \hat{b}_0}{\exp(\lambda \Delta) - 1} \text{ and } \hat{b}_0 = \frac{\bar{Y}(\exp(2\lambda \Delta) - 1)}{S^2 (\exp(\lambda \Delta) - 1)}.$$

On the other side, by Proposition 3.9, the characteristic function of the Lévy functional equals:

$$\begin{aligned}
\phi(\vartheta) &= \exp\left(\lambda \int_0^{\Delta} a\left(\frac{b}{b - i\vartheta \exp(\lambda s)} - 1\right) ds\right) = \exp\left(\lambda \int_0^{\Delta} a\left(\frac{b}{b - i\vartheta \exp(\lambda s)} \frac{b + i\vartheta \exp(\lambda s)}{b + i\vartheta \exp(\lambda s)} - 1\right) ds\right) \\
&= \exp\left(a\lambda b \left(\int_0^{\Delta} \frac{b}{b^2 + \vartheta^2 \exp(2\lambda s)} ds + i \int_0^{\Delta} \frac{\vartheta \exp(\lambda s)}{b^2 + \vartheta^2 \exp(2\lambda s)} ds\right) - a\lambda \Delta\right) \\
&= \exp\left(a \log\left(\frac{a - i\vartheta}{b - i\vartheta \exp(\lambda \Delta)}\right)\right) = \left(\frac{b - i\vartheta}{b - i\vartheta \exp(\lambda \Delta)}\right)^a.
\end{aligned}$$

Note that a peculiar property of this process emerges by observing that for any k, $\frac{y_k + x_{k-1}}{x_k} = \exp(\lambda \Delta)$, where we used the observed values of Y_k *and* $X(k\Delta)$ of the expression in (16). This way, if the $\Gamma(a, b)$-OU process does not jump between consecutive periods $(k-1)\Delta$ *and* $k\Delta$, then λ can be exactly recovered by

$$\lambda = \frac{1}{\Delta} \log\left(\frac{x_{k-1}}{x_k}\right).$$

4.2 Inference for a $TS(\kappa, a, b)$-OU process

Now we have 3 parameters to estimate, so we need to solve a 3 variables system using mean, variance and bias to find the initial estimators.

$$\mathbb{E}[Y_1] = 2a\kappa b^{\frac{\kappa-1}{\kappa}}(\exp(\lambda\Delta - 1)), \quad Var(Y_1) = 4a\kappa(1-\kappa)b^{\frac{\kappa-2}{\kappa}}(\exp(2\lambda\Delta) - 1),$$

$$B(Y_1) = \frac{\mathbb{E}[(Y_1 - \mathbb{E}[Y_1])^3]}{Var(Y_1)^{\frac{3}{2}}} = \frac{(2-\kappa)(\exp(3\lambda\Delta) - 1)}{\sqrt{ab\kappa(1-\kappa)}(\exp(2\lambda\Delta)^{\frac{3}{2}})}.$$

Then, we obtain the moment-based initial estimator $\hat{\theta}_0 = (\hat{\kappa}_0, \hat{a}_0, \hat{b}_0)$, where

$$\hat{\kappa}_0 = \frac{B\bar{Y}(\exp(2\lambda\Delta) - 1)^2 - 2S(\exp(\lambda\Delta) - 1)(\exp(3\lambda\Delta) - 1)}{B\bar{Y}(\exp(2\lambda\Delta) - 1)^2 - S(\exp(\lambda\Delta) - 1)(\exp(3\lambda\Delta) - 1)},$$

$$\hat{b}_0 = \left(\frac{2\bar{Y}(1-\hat{\kappa}_0)(\exp(2\lambda\Delta) - 1)}{S^2(\exp(\lambda\Delta) - 1)}\right)^{\hat{\kappa}_0}, \text{ and } \hat{a}_0 = \frac{\bar{Y}\hat{b}_0^{\frac{1-\hat{\kappa}_0}{\hat{\kappa}_0}}}{2\hat{\kappa}_0(\exp(\lambda\Delta) - 1)}.$$

We clearly have that $\hat{\kappa}_0 < 1$, but no guarantee is given that $\hat{\kappa}_0$ will remain strictly positive. To cut out the possibility that $\hat{\kappa}_0 \leq 0$ we could restrict the tempered stable distribution to the case where $\kappa = \frac{1}{2}$ and we would be working with an inverse Gaussian distribution. Another alternative is to use the generalized method of moments.

On the other side, according to Proposition 3.9, the characteristic function of the Lévy functional $Z^*(\Delta)$is given by

$$\begin{aligned}
\phi(\vartheta) &= \exp\left(\lambda\int_0^\Delta C_{Z(1)}(\vartheta\exp(\lambda s))\,ds\right) \\
&= \exp\left(\lambda\int_0^\Delta 2i\vartheta\exp(\lambda s)a\kappa(b^{\frac{1}{\kappa}} - 2i\vartheta\exp(\lambda s))^{\kappa-1}\,ds\right) \\
&= \exp\left(a\left((b^{\frac{1}{\kappa}} - 2i\vartheta)^\kappa - (b^{\frac{1}{\kappa}} - 2i\vartheta\exp(\lambda\Delta))^\kappa\right)\right).
\end{aligned}$$

4.3 Inference for a $NIG(\alpha, \beta, \delta)$-OU process

Similarly to the tempered stable case, we need the empirical mean, variance and bias for the initial estimators of the parameters. We have:

$$\mathbb{E}[Y_1] = \frac{\delta\beta\exp(\lambda\Delta) - 1}{\sqrt{\alpha^2 - \beta^2}}, \quad Var(Y_1) = \frac{\alpha^2\delta\exp(2\lambda\Delta) - 1}{\sqrt{\alpha^2 - \beta^2}}, \quad B(Y_1) = \frac{3\beta\exp(3\lambda\Delta) - 1}{\alpha\sqrt{\delta(\alpha^2 - \beta^2)^{\frac{1}{2}}(\exp(2\lambda\Delta) - 1)^{\frac{3}{2}}}},$$

and solving the system we obtain resulting moment-based initial estimator $\hat{\theta}_0 = (\hat{\alpha}_0, \hat{\beta}_0, \hat{\delta}_0)$:

$$\hat{\alpha_0} = \frac{(\bar{Y^2} + (\exp(\lambda\Delta) - 1)\hat{\delta}_0^2)^{\frac{3}{2}}(\exp(2\lambda\Delta) - 1)}{\hat{\delta} S^2 (\exp(\lambda\Delta) - 1)^3},$$

$$\hat{\beta}_0 = \frac{\hat{\alpha}_0 \bar{Y}}{\sqrt{\bar{Y^2} + ((\exp(\lambda\Delta) - 1)\hat{\delta}_0)^2}},$$

and

$$\hat{\delta}_0 = \sqrt{\frac{3S\bar{Y}(\exp(3\lambda\Delta) - 1)}{B(\exp(2\lambda\Delta) - 1)^2(\exp(\lambda\Delta) - 1)} - (\frac{\bar{Y}}{\exp(\lambda\Delta) - 1})^2}.$$

If it would happen for the last radical to be negative one could use the generalized method of moments. Another strategy would be to turn to the symmetric case ($\beta = 0$) and obtain the estimator by matching the theoretical and empirical kurtosis:

$$\frac{\mathbb{E}[(\,Y_1 - \mathbb{E}[\,Y_1])]^4}{Var(\,Y_1)^2} = \left(1 + \frac{(\exp(2\lambda\Delta) + 1)}{\alpha\delta(\exp(2\lambda\Delta) - 1)}\right),$$

which leads to the initial estimator:

$$\hat{\alpha}_0 = \frac{\hat{\delta}_0(\exp(2\lambda\Delta) - 1)}{S^2}, \quad \hat{\delta}_0 = \frac{S}{(\exp(2\lambda\Delta) - 1)}\sqrt{\frac{3(\exp(2\lambda\Delta) + 1)}{K - 3}}.$$

Finally, the characteristic function of the Lévy functional $Z^*(\Delta)$is given by:

$$\begin{aligned}
\phi(\vartheta) &= \exp\left(\lambda \int_0^{\Delta} \frac{\delta i\vartheta \exp(\lambda s)(\beta + i\vartheta \exp(\lambda s))}{\alpha^2 - (\beta + i\vartheta \exp(\lambda s))^2} ds\right) \\
&= \exp\left(\delta \int_{\vartheta}^{\vartheta \exp(\lambda\Delta)} \frac{i(\beta + iy)}{\sqrt{\alpha^2 - (\beta + iy)^2}} dy\right) \\
&= \exp\left(\delta(\sqrt{\alpha^2 - (\beta + i\vartheta)^2} - \sqrt{\alpha^2 - (\beta + i\vartheta \exp(\lambda\Delta))^2})\right).
\end{aligned}$$

5. On the simulation of a D-OU process

Following [20] and adding further research, we work on obtaining the simulations of Ornstein-Uhlenbeck processes. A general method to simulate a D-OU process is based on the recursive expression of X_t and the Euler scheme. We will discretize the time line with equidistant nodes separated by Δ, and this way we have that $X(t_k) = X(k\Delta)$ and thus:

$$X(k\Delta) = \exp(-\lambda\Delta)\left(X((k-1)\Delta) + \sum_{j=1}^{[\frac{\lambda\Delta}{\bar{h}}]} \exp(j\bar{h})Z(\bar{h})\right), \tag{17}$$

where $k = 1, 2, \dots, n$, $0 \leq \bar{h} \leq \lambda\Delta$ is sufficiently small, $[x]$ denotes the maximum integer of x and $X(0)$ is simulated from the stationary distribution D. This scheme is justified by the following result.

Proposition 5.1.

$$\sum_{j=1}^{[\frac{\lambda\Delta}{\bar{h}}]} \exp(j\bar{h})Z(\bar{h}) \xrightarrow{d} \int_0^{\lambda\Delta} \exp(s)dZ(s) \quad \textit{as} \quad \bar{h} \to 0,$$

where $\xrightarrow{d}$ denotes convergence in distribution.

Proof [20]: Let $n = [\frac{\lambda\Delta}{\bar{h}}]$ be the maximum integer of $\frac{\lambda\Delta}{\bar{h}}$ and consider the partition $s_k = k\bar{h}$ with $k = 0, 1, 2, \dots, n+1$. First of all we have this process that converges uniformly to the exponential function $\exp(s)$:

$$H_n(s) = \mathbf{1}_{\{0\}}(s) + \sum_{k=1}^{n+1} \exp((k-1)\bar{h})\mathbf{1}_{](k-1)\bar{h}, k\bar{h}]}(s)$$

as $H_n(0) = 1$ and for $s_k \geq 1$, $H_n(s) \overset{n\to\infty}{\to} \exp(s)$. Then, according to [12], Z is a semimartingale and it follows that as $\bar{h} \to 0$, $\int_0^{\lambda\Delta} H_n(s)dZ(s)$ converges in probability to $\int_0^{\lambda\Delta} \exp(s)dZ(s)$. Finally,

$$\begin{aligned}\int_0^{\lambda\Delta} H_n(s)dZ(s) &= \sum_{k=1}^{n+1} \exp((k+1)\bar{h})(Z(k\bar{h}) - Z((k-1)\bar{h})) \\ &= \sum_{j=1}^{n} \exp(j\bar{h})(Z((j+1)\bar{h}) - Z(j\bar{h})) \overset{d}{=} \sum_{j=1}^{n} \exp(j\bar{h})Z(\bar{h}).\end{aligned}$$

The first equality follows from the definition of H_n and the geometric interpretation of the integral, as we can think the usual idea of the area of the rectangle with height $H_n(s)$ and, in this case, the base depends on the values of the Lévy process Z at the extremes of the intervals. Then we just apply a translation from k to $j+1$ and use the linearity of Z. □

Now we start in detail the discussion of the D-OU process simulation.

5.1 Simulation of a $\Gamma(a, b)$-OU process

As mentioned before, the BDLP $Z = \{Z(t)\}$ of this process is a compound Poisson process with intensity parameter a and associated exponential distribution with parameter b. Hence, if we define $\mathbf{N} = \{N(t)\}$

as a Poisson process with intensity parameter a and $Y_1, Y_2, ...$ as independent and identically distributed exponential random variables with parameter b, we can represent the Lévy process in law as

$$Z(t) = \sum_{j=1}^{N(t)} Y_j,$$

Using this representation, it is not difficult to see that the Lévy functional $Z^*(\Delta) = \int_0^{\lambda\Delta} \exp(s) dZ(s)$ is explicitly given by:

$$Z^*(\Delta) \stackrel{d}{=} \sum_{j=1}^{N(\lambda\Delta)} \exp\left(T_j\right) Y_j,$$

by thinking again about the geometric interpretation of the integral, where $T_1 < T_2 < T_3 ...$ are the arrival times of the Poisson process $\mathbf{N}$. As a consequence, we obtain the following exact simulation scheme:

$$X(k\Delta) = \exp(-\lambda\Delta)\left(X((k-1)\Delta) + \sum_{j=1}^{N(\lambda\Delta)} \exp(T_j)\, Y_j\right), \tag{18}$$

where $X(0)$ can be simulated from a $\Gamma(a, b)$ distribution.

First of all we need to estimate the sample path of a Poisson process. As we need the time discretization to be under control in order to have both Poisson process and Gamma-OU in the same time grid, we will use the exponential spacings method. As we have already stated in the first chapter, with the uniform method estimation we can easily get the arrival times in the interval $[0, T]$ and also faster, but to make our code easier, we go with the exponential spacings method, thus obtaining vectors of the same length for both the OU and the Poisson processes, with each element indexing a point in the grid.
Then, we just apply a similar expression for the OU process estimation thanks to Proposition 3.5:

$$X(k\Delta) = \exp(-\lambda\Delta)X((k-1)\Delta) + \sum_{n=N_{(n-1)\Delta}+1}^{N_{n\Delta}} Y_n, \tag{19}$$

where Y_i are, again, i.i.d. exponential random variables with parameter b. The simulation algorithm would be

1. Generate a Poisson process sample path using the exponential spacings method.
2. Generate N exponential random values with parameter b.
3. Draw a $\Gamma(a, b)$ random value for X_0.
4. Apply the scheme in (19), with the convention that if the lower bound is equal or higher that the upper bound, the sum is set to 0.

A sample plot obtained with parameters $a = 10$, $b = 100$, $X_0 = 0.08$, $T = 1$ and a discretization of $N = 2^{12}$ intervals would look like the one in Figure 7.

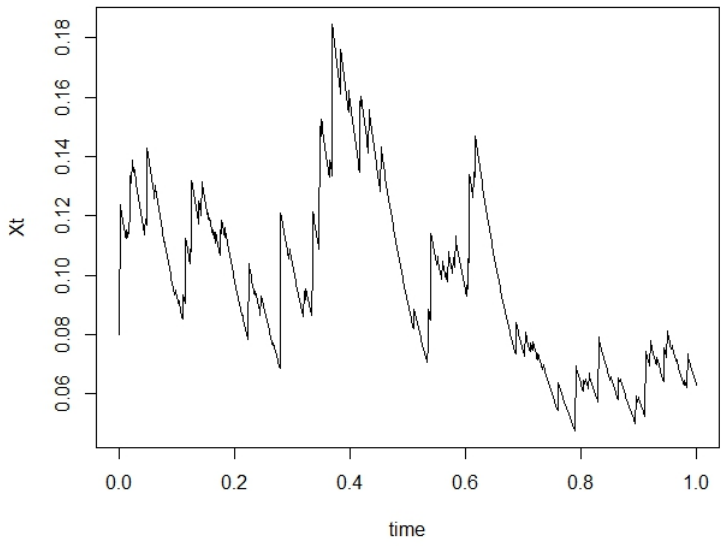

Figure 7: Simulation of a gamma-OU process

5.2 Simulation of a $TS(\kappa, a, b)$-OU process

As we mentioned before, the BDLP $Z = \{Z(t)\}$ of this process can be decomposed in law as:

$$Z(t) = I(t) + \sum_{j=1}^{N(t)} Y_j,$$

where $\mathbf{I} = \{I(t)\}$ is a $TS(\kappa, \kappa a, b)$-Lévy process, $\mathbf{N} = \{N(t)\}$ is a Poisson process with intensity parameter $ab\kappa$ and $Y_1, Y_2, ...$ are independent and identically distributed $\Gamma(1-\kappa, \frac{b^{\frac{1}{\kappa}}}{2})$ random variables.
Hence, discretizing again the time in a uniform grid we can use the following scheme:

$$X(k\Delta) = \exp(-\lambda\Delta)\left(X((k-1)\Delta) + \sum_{j=1}^{[\frac{\lambda\Delta}{\bar{h}}]} \exp(j\bar{h})I(\bar{h}) + \sum_{j=1}^{N(\lambda\Delta)} \exp(T_j)\, Y_j\right),$$

where $T_1 < T_2 < T_3 ...$ are the arrival times of the Poisson process $\mathbf{N}$, $\bar{h}$ is sufficiently small and $X(0)$ is simulated from a $TS(\kappa, a, b)$ distribution. If we had $\kappa = \frac{1}{2}$ we could easily simulate $I(\bar{h})$ as it would be an inverse Gaussian random variable. However, as we have already commented, the density function of a general tempered stable random variable $I(\bar{h})$ is not explicitly known and its series representation has no practical use. Thus, we need to invoke the method proposed by Rosinski [**?**],[**?**]. He shows that the process $\{J_M(t)\}_{t\in[0,T]}$ with

$$J_M(t) = 2\sum_{j=1}^{M} \min\left\{\left(\frac{a\lambda T\kappa}{b_j\Gamma(1-\kappa)}\right)^{\frac{1}{\kappa}}, \frac{e_j v_j^{\frac{1}{\kappa}}}{b^{\frac{1}{\kappa}}}\right\} \exp(\lambda T u_j)\mathbf{1}_{Tu_j \le t}$$

converges almost surely and uniformly as $M \to \infty$ to the process $\left\{J(t) = \int_0^{\lambda t} \exp(s) dI(s)\right\}_{t\in[0,T]}$. Here,$\{e_j\}$ is an *i.i.d.* sequence of exponential random variables with parameter 1, $\{u_j\}, \{v_j\}$ are *i.i.d.* sequences of uniform random variables on the interval [0,1] and $b_1 < b_2 < b_3, ...$ are the arrival times of a Poisson process with intensity parameter 1. All series are assumed to be independent of each other. This method permits a simulation of the whole path of the process $\{J(t)\}_{t\in[0,T]}$. As we have already simulated a TS process, we can generate one random value from there to have the random variable X_0.

As a result, we can rely on the following simulation scheme:

$$X(k\Delta) = \exp(-\lambda k\Delta)\left(X(0) + J_M(k\Delta) + \sum_{j=1}^{N(\lambda k\Delta)} \exp(T_j)\, Y_j\right),$$

where M is sufficiently large, $T = n\Delta$ and $X(0)$ can be simulated as stated before.

Note that not only we are changing the part of the I approximation, but also we no longer use the recursive scheme. We can do that thanks to Proposition 3.5, as it lets us calculate the new grid interval of the Lévy functional either using the information on the past node or from all nodes starting from zero, obtaining the same results in law.

5.3 Simulation of a $NIG(\alpha, \beta, \delta)$-OU

This section will mainly be based on the work of E.Taufer and N.Leonenko in [19], trying to focus overall in the simulation part rather than the deep theory behind it. When looking for literature related to the simulation of Lévy driven OU processes, one mainly finds two ways to simulate the paths. The first one would be the one we applied already, based on the simulation of the sample paths of the BDLP Z and then approximating the corresponding integral by way of sums. Despite the lack of a closed form expression for the density of the TS process, this approach has been useful for both the TS and the Gamma-OU. However, depending on the nature of Z, this may be a little more complicated.

This would be the case of the NIG-OU, as its BDLP is composed of three independent Lévy processes that need to be separately simulated, and one needs to take special care for the very small jumps, as already referenced from Schoutens [17]. The second way to simulate the paths may be preferable in this case, where we will simulate directly the error term of:

$$X_j = e^{-\lambda} X_{j-1} + \epsilon_j, \text{ with } \epsilon_j \overset{d}{=} e^{-\lambda} \int_0^1 e^{-\lambda s} dZ(\lambda s), \tag{20}$$

via series representation, rather than first simulating from the BDLP. This procedure requires the inversion of the tail mass function of the Lévy measure of the BDLP but explicit expressions are rarely available and in some cases one needs to resort to either analytical or numerical approximations. A second drawback that may lay here is the slow convergence of the series involved.

For the NIG process it turns out simpler to apply this procedure rather than simulating three Lévy processes. The characteristic function needed to simulate an OU process with marginal $NIG(\alpha, \beta, \delta)$ and parameter λ has the form

$$\phi_t(\vartheta) = \exp\left(-\delta(\sqrt{\alpha^2 - (\beta + i\vartheta)^2}) - \sqrt{\alpha^2 - (\beta + i\vartheta t \exp(-\lambda))^2}\right).$$

where we have considered the location of the process be at $\mu = 0$ and suppose we observe the process at time instants $t = 0, 1, ...$, i.e. $t_i - t_{i-1} = \Delta = 1$. Then, the usual approach in numerically estimating a

distribution function $F(x)$ is to numerically evaluate the Gil-Palaez formula

$$F(x) = \frac{1}{2} - \int_{-\infty}^{\infty} \frac{\phi(\vartheta)}{2\pi i\vartheta} e^{ix\vartheta} d\vartheta. \tag{21}$$

In our case, we will need to determine sample intervals and truncation limits necessary for a given accuracy directly from the analytical form of the characteristic function. Inversion will be over the interval $[-D/2, D/2]$. Following [8], the inversion can be defined by

$$F^N(x) = \sum_{k=-(N/2)}^{N/2} G[k] e^{i2\pi kx/T}, \tag{22}$$

being $F^N(x)$ the numerical approximation of $F(x)$ and the coefficients $G[k]$ given by

$$G[k] = \begin{cases} \frac{1}{2}, & \text{for } k = 0, \\ \frac{1-\cos(2\pi k\eta)}{ik2\pi}\phi(-2\pi k/T), & \text{for } 0 < |k| < N/2, \\ 0, & \text{for } |k| = N/2. \end{cases} \tag{23}$$

where $\eta = D/T$ is the fraction of the interval $(-T/2, T/2)$ over which $F^N(x)$ accurately approximates $F(x)$. This result, also extracted from [8], gives us conditions under which for any given $\epsilon > 0$ and $D > 0$, there exist values of η, T and N such that $\left|F(x) - F^N(x)\right| \leq \epsilon$ for every $|x| \leq D/2$.

Lemma 5.2. *Let X a random variable with $\mathbb{E}(X^2) < \infty$ with distribution $F(x)$ and characteristic function $\phi(\vartheta)$. Suppose that*

1. *There exist constants A and $a > 1$ such that $F(-x) \leq A\,|x|^{-a}$ and $1 - F(x) \leq A\,|x|^{-a}$.*
2. *There exist constants B and b such that $|\phi(\vartheta)| \leq B\,|\vartheta/2\pi|^{-b}$ for all real ϑ.*

Then, for constants $0 < \eta < 2/3$, $T > 0$ and $N > 0$, on the interval $|x| \leq \eta T/2$,

$$\left|F(x) - F^N(x)\right| \leq AT^{-a}L_1(\eta, a) + \frac{2BT^b}{\pi} z(b+1, N/2),$$

where $z(u, v) = \sum_{k=0}^{\infty}(k+v)^{-u}$ is the Hurwitz zeta function and

$$L_1 = (\frac{\eta}{2})^{-a} + 2z(a, 1 - \frac{1}{2}\eta) + z(a, 1 - \frac{3}{2}\eta).$$

Moreover, for any $D > 0$ and $\epsilon > 0$, choosing $\eta \in (0, 2/3)$, T and N such that

$$\eta^a L_1(\eta, a) \leq 2^{a+1}, \quad T \geq \max\left(\frac{D}{\eta}, \frac{2}{\eta}\left(\frac{3A}{\epsilon}\right)^{1/a}\right), \quad N \geq 2 + 2T\left(\frac{6B}{\epsilon\pi b}\right)^{1/b}, \tag{24}$$

suffices to guarantee that $\left|F(x) - F^N(x)\right| \leq \epsilon$ for every $|x| \leq D/2$. It is always possible to choose η to meet the above conditions.

As we do not know $F(x)$, only the characteristic function of the process, the parameters A and a must be determined from the characteristic function alone. For this, we resort to the following lemma:

Lemma 5.3. *For a distribution function $F(x)$ with characteristic function $\phi(\vartheta)$, for any non-negative integer n and any $x \geq 0$,*

$$F(-x) \leq A\,|x|^{-2n} \quad \textit{and} \quad 1 - F(x) \leq A\,|x|^{-2n}\,, \tag{25}$$

where $A = \mathbb{E}(X^{-2n}) = (-1)^n \phi^{(2n)}(0)$, provided the derivative actually exists.

Given that we have an array of numbers (data) of length N, $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}$, the *Discrete Fourier Transform (DFT)* of x is a vector $\hat{x} \in \mathbb{R}^n$ defined by

$$\hat{x}_m = \sum_{j=1}^{N} x_j w_N{}^{(j-1)(m-1)},$$

where $w_N = e^{-\frac{2\pi}{N}}$, N root of 1. The FFT is an algorithm that allows the calculation of the sum with cost $\mathcal{O}(N \log N)$ instead of $\mathcal{O}(N^2)$. Note that this approximation of the distribution function does not directly depend on A, a, B and b, which are needed to determine the values of $\eta \in (0, 2/3)$, T and N for given ϵ and D. N can be any number, however when tabulating the whole distribution it should be a power of 2 so that the FFT can be used.

Regarding the choice of these last values, since the computational effort, as we said, is proportional to $N \log N$ using the FFT, it is preferable to choose η as large as possible and T small. Here there is a table (Table 2) extracted from Hughett (1998) [8] that provides the optimal values of η for given a, the only parameter on which depends the choice of η.

a	1.125	1.25	1.5	2	3	4	5	10
η_{opt}	0.0855	0.1874	0.3530	0.4666	0.4955	0.4990	0.4997	0.5000

Table 2: Optimal values of η for given a.

Hughett does not only gives us the scheme to follow in order to obtain the values of $F^N(x)$, but also some consistency checks to detect errors in the distribution:

i) The minimum and maximum values of $F^N(x)$ should be within ϵ of 0 and 1, as it is a cumulative distribution function.

ii) Values should be approximately monotone, more precisely, the smallest increment $F^N(x_{n+1}) - F^N(x_n)$ should not be less than -2ϵ.

After this bit of theory we can get back to our case. We consider a $NIG(2, 1, 1)$ distribution and we choose $\lambda = 1$. We fix the error in approximation as $\epsilon = 0.0001$ and proceed to determine the values of T and N needed for the approximation of the Gil-Palaez formula (21). To determine a bound on T, we need the values of a, A and η, as one can see from (24), and there is no need of choosing $D < 2(3A/\epsilon)^{1/a}$, as it will not reduce the minimum acceptable value of T. Then, choosing η accordingly to Table 2, we obtain the following values of T:

Now we need to determine B and b in order to get the minimum value of N, following again the expressions in (24). Following Lemma 5.2, for a choice of b, it holds that $B = \max_\vartheta \left|\vartheta^b \phi(\vartheta)\right|$. We can compute numerically these values, and using $a = 5$, we obtain the values for N shown in Table 4.

As we needed values of N to be a power of 2, we can use $N = 2^{14} = 16384$. Clearly, if we parametrize the distribution differently, we will need other values of N and T.

a	2	3	4	5
$T \geq$	1357.69	492.65	368.67	346.27

Table 3: Values of T for different values of a.

b	4	6	8
$N \geq$	14729	9753	8789

Table 4: Values of N for different values of b, with $a = 5$.

So now we have the back-end ready for our simulation. On the one hand we have the relation (20) and numerical values of the cumulative function of the distribution followed by $\epsilon_j, j = 1, \dots, n$. What we lack now is an *i.i.d.* sequence of these values. For doing so, we will use the **Inverse transform sampling** or "golden rule", which is a basic method for pseudo-random number sampling: a method for generating sample numbers at random from any probability distribution given its cumulative distribution function. It goes as follows:

Let X be a random variable whose distribution is described by the cumulative function $F(x)$, then

1. Generate a random number $u \sim U(0, 1)$.
2. Compute the value x such that $F(x) = u$.
3. Take x to be the random number drawn from the distribution described by F.

As the data we have from the cumulative function $F(x)$ is discrete, we either need a really dense grid or interpolate the points in order to obtain the right value.

Summing it all up, the simulation scheme would have the following steps:

1. Choose the process' parameters and obtain N, T and η accordingly. For example, for a NIG(2,1,1) we used $N = 16384$, $T = 346.27$ and $\eta = 0.4997$.
2. Generate the function $G[k]$ from (23).
3. Apply the inversion formula (22). Now we have numerically the cumulative function $F(x)$.
4. Use the golden rule to draw *i.i.d.* values from the distribution the numerical values obtained before.
5. Obtain the NIG-OU values by applying the relation 20.

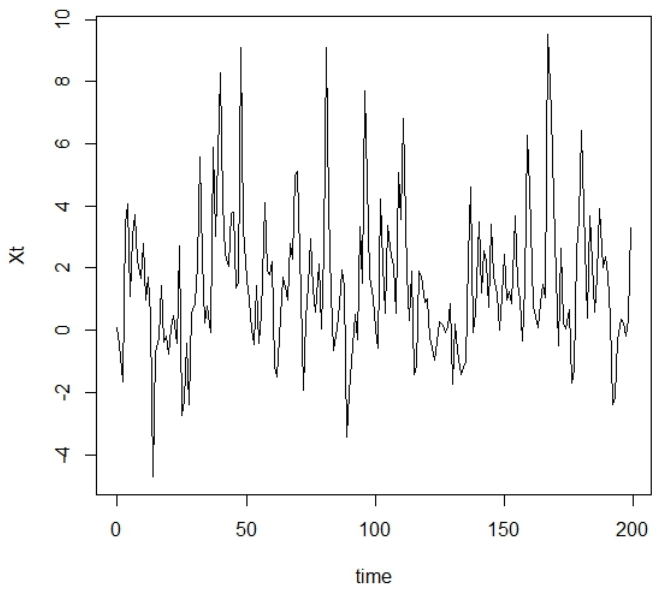

Figure 8: Simulation of an NIG(2,1,1)-OU sample path with $n = 200$ time points.

6. Maximum Empirical Likelihood

In section 4 we introduced the planning for the likelihood estimation of the parameters. One of the big problems encountered there was the need of the density function, whose expression is not known, so it had to be approximated via the inverse Fourier transform applied to the characteristic function of the Lévy functional, which can easily lead to numerical errors. Another issue faced when solving the maximization of the likelihood functions was that we were working with complex numbers, and optimization packages usually do not treat them well. In order to overcome this, we explore a new estimation method proposed by Qin & Lawless (1994) [13] based on the Empirical Likelihood specially applied to the estimation of Lévy processes, and more generally infinitely divisible distributions.

Thus, in this section we will introduce the derivation of the method just to Lévy processes such as the gamma, tempered-stable and inverse Gaussian (from which we already know the characteristic function) following the work of [7]. We can not apply it directly to the Lévy-OU processes as we do not know the expression of its characteristic function.

As we have just mentioned, the distribution functions of infinitely divisible distributions do not always have a closed form, so this method relies on the empirical characteristic function instead. The goal is then to estimate the vector of parameters θ for the population with given a set of $(y_1, y_2, \dots, y_n)$ iid. observations. The *empirical characteristic function* is defined as:

$$\phi_n(u) = \frac{1}{n} \sum_{i=1}^{n} e^{iuy_j},$$

where $u \in \mathbb{R}$. Then as the observations are iid. we can deduce the following:

$$\mathbb{E}[\phi_n(u)] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[e^{iuy_j}] = \mathbb{E}[e^{iuy_j}] = \phi_\theta(u),$$

where $\phi_\theta(u)$ is the theoretical characteristic function of the distribution Y_j of the observation y_j. Using this relation, we can construct an unbiased estimation function

$$\mathbf{h}(u, y_j, \theta) = e^{iuy_j} - \phi_\theta(u).$$

Indeed, we have a vector-valued function $\mathbf{h}(\theta)$ which is unbiased, meaning that

$$\mathbb{E}[\mathbf{h}(u, y_j, \theta)] = 0 \ \ \forall u \in \mathbb{R}.$$

Invoking now the Euler's formula it provides that $e^{iuy_j} = cos(uy_j) + isin(uy_j)$, as we are interested in separating the real and imaginary part, obtaining

$$\mathbf{h}(u, y_j, \theta) = \cos(uy_j) - \phi_\theta^R(u) + i(\sin(uy_j) - \phi_\theta^I(u)),$$

where $\phi_\theta^R(u)$ and $\phi_\theta^I(u)$ are respectively the real and imaginary part of the characteristic function $\phi_\theta(u)$. To obtain the estimates we will plug in k values of u and exploit the equation above, thus getting a system of $2k$ equations to be solved, k for the real part and k for the imaginary part. So let $\mathbf{u} = (u_1, \dots, u_k)^T$, we look at a $2k \times 1$ vector

$$\mathbf{h}(y_j, \theta) = (\mathbf{h}^R(y_j, \theta), \mathbf{h}^I(y_j, \theta))^T, \tag{26}$$

where

$$\mathbf{h}^R(y_j, \theta) = (\cos(u_1 y_j) - \phi_\theta^R(u_1), \dots, \cos(u_k y_j) - \phi_\theta^R(u_k))^T,$$
$$\mathbf{h}^I(y_j, \theta) = (\sin(u_1 y_j) - \phi_\theta^I(u_1), \dots, \sin(u_k y_j) - \phi_\theta^I(u_k))^T.$$

Keep in mind that we do know the theoretical characteristic functions of the Lévy processes, which depend on the unknown vector of parameters θ. With this preliminary part done, we can start deriving the estimation method. The idea is to keep on arranging expressions with elements that remain unknown for us and obtain a final formula that only depends on θ, which we will then maximize to get the estimates.

Definition 6.1. The empirical likelihood function is defined as, given the observations $y_1, \dots, y_n$ of iid. random variables $Y_1, \dots, Y_n$, is

$$L(F_\theta) = \begin{cases} \prod_{i=1}^n f_\theta(y_i) \text{ if } Y_i \text{ are from a continuous distribution}, \\ \prod_{i=1}^n \mathbb{P}(Y_i = y_i) \text{ if } Y_i \text{ are from a discrete distribution}, \end{cases}$$

where f_θ is the density function of the Y_i in the continuous case, which is the one we are interested in.

We will denote f_θ as p_i, so the right-hand side of the likelihood function becomes $\prod_{i=1}^n p_i$. This expression is maximized by the empirical distribution function

$$F_n(y) = \frac{1}{n}\{x : x \in \{x_1, \dots, x_n\}, x \le y\} \ = \ \frac{1}{n}\sum_{i=1}^{n} I_i,$$

that is the number of observations in the sample which are less than or equal to y, divided by n to get a distribution, and I_i is the indicator function for $x_i \le y$.

Definition 6.2. The empirical likelihood ratio, for a distribution F with density function f when the underlying distribution is continuous, or probability function $p_i = \mathbb{P}(Y_i) = i$ for the discrete case, is

$$R(F) = \frac{L(F)}{L(F_n)}.$$

Since the likelihood function for the empirical distribution is $\frac{1}{n}, \forall n$, as all possible orders of the set are equally likely, we obtain

$$R(F) = \prod_{i=1}^{n} np_i,$$

and if we take the log

$$l_n(\theta) = \log(R(F)) = \log \prod_{i=1}^{n} np_i = \sum_{i=1}^{n} \log np_i.$$

Definition 6.3. The maximum empirical likelihood estimator (*MELE*) is the choice of F_θ which maximizes $R(F_\theta)$ subject to the constraints

$$p_i \geq 0, \sum_{i}^{n} p_i = 1, \sum_{i}^{n} p_i \mathbf{h}(y_i, \theta) = 0.$$

The first two constraints are self-explanatory, arising from probability properties. The last one is the empirical representation of the unbiased equation introduced previously.

Now we find ourselves facing a problem of maximization under constraints, which leads us directly to solving the following Lagrangian function:

$$L_n(\theta) = \sum_{i=1}^{n} log np_i - \mu \left[\sum_{i}^{n} p_i - 1 \right] - n\eta^T \left[\sum_{i}^{n} p_i \mathbf{h}(y_i, \theta) \right],$$

where μ and $\eta^T = (\eta_{1,1}, \dots, \eta_{1,k}, \eta_{2,1}, \dots, \eta_{2,k})^T$ are Lagrange multipliers. Now if we take the derivative with respect to p_i and set equal to zero we obtain the following equality:

$$\frac{\partial Ln(\theta)}{\partial p_i} = \frac{n}{np_i} - \mu - n\eta^T \mathbf{h}(y_i, \theta) = 0,$$

$$p_i = \frac{1}{\mu + n\eta^T \mathbf{h}(y_i, \theta)}.$$

Now multiplying by p_i we get:

$$p_i \frac{\partial Ln(\theta)}{\partial p_i} = 1 - p_i\mu - p_i n\eta^T \mathbf{h}(y_i, \theta) = 0, \ \forall i.$$

So if we now sum by all p_i

$$\sum_{i=1}^{n} p_i \frac{\partial Ln(\theta)}{\partial p_i} = \sum_{i=1}^{n} (1 - p_i\mu - p_i n\eta^T \mathbf{h}(y_i, \theta)) =$$

$$n - \mu \sum_{i=1}^{n} p_i - n\eta^T \sum_{i=1}^{n} p_i \mathbf{h}(y_i, \theta) =$$
$$n - \mu - n\eta^T \mathbf{0} = 0$$
$$\iff n = \mu.$$

And we can rewrite the expression of p_i as:

$$p_i = \frac{1}{n} \frac{1}{1 + n\eta^T \mathbf{h}(y_i, \theta)}.$$

Using the third restriction with these values of p_i provides a system for calculating $\eta = \eta(\theta)$:

$$\mathbf{0} = \sum_{i}^{n} p_i \mathbf{h}(y_i, \theta) = \frac{1}{n} \sum_{i}^{n} \frac{1}{1 + \eta^T \mathbf{h}(y_i, \theta)} \mathbf{h}(y_i, \theta). \tag{27}$$

That can be shown to have a unique solution.

Substituting this result into the empirical log-likelihood ratio function

$$l_n(\theta) = \sum_{i=1}^{n} log \left(n \frac{1}{n} \frac{1}{1 + \eta^T \mathbf{h}(y_i, \theta)} \right) = \sum_{i=1}^{n} \log((1 + \eta^T \mathbf{h}(y_i, \theta))^{-1}).$$

Theorem 6.4. *The MELE is obtained by maximizing*

$$l_n(\theta) = \sum_{i=1}^{n} \log((1 + \eta^T \mathbf{h}(y_i, \theta))^{-1}) = -\sum_{i=1}^{n} \log(1 + \eta^T \mathbf{h}(y_i, \theta))$$

subject to the $2k \times 1$ *restrictions of (26) and* $\hat{\eta}$ *the solution to (27). We then define*

$$\hat{\theta} = \arg\max l_n(\theta).$$

N.B: We worked on implementing the MELE algorithm to test the estimates against the ones from MLE. Both software R and Matlab yielded no usable results due to package optimization limits. Then, we tried with Julia and the open-source optimization package IPOPT. We contacted an expert which told us that due to the restriction (27), where η is a vector, the solving is considerably difficult and the problem would need a reformulation.

The other option would have been working directly with the IPOPT package without the Julia user interface (JuMP), but it would require to input both jacobian and hessian matrices for the constraint, which does not seem like a better approach at all. We keep the problem reformulation and its implementation as a future work.

References

[1] A. Arratia, A. Cabaña, E.M. Cabaña *A construction of continuous-time ARMA models by iterations of Ornstein Uhlenbeck processes*, SORT, v40(2), 267-302, 2016.

[2] A. Arratia, *Computational finance, an introductory course with R*, Springer, 2014.

[3] Barndorff-Nielsen O. E. , *Processes of normal inverse Gaussian type*, Finance and Stochastic 2, 41-68, 1998.

[4] Barndorff-Nielsen O. E., N. Shepard, *Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics*, Journal of the Royal Statistical Society B, 63, 167–241, 2001.

[5] Barndorff-Nielsen O. E., Shepard N *Normal modified stable processes*, Theory ProbabMath Stat, 2:1–20 2002.

[6] C. Halgreen, *Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions*, Zeitschrift für Warrscheinlichkeitstheorie un verwandte Gabiete 47, 13-18, 1979.

[7] Henry Elgin, University of Essex, MSc Dissertation, 2011.

[8] P. Hughett, *Error bounds for numerical inversion of a probability characteristic function*, SIAM J. Numer. Anal. 34, no. 4, 1368-1392. 1998.

[9] I. Karatzas, S.E. Shreve, *Brownian Motion and Stochastic Calculus, second ed*, Springer, New York, 1991.

[10] Prof. Thomas G. Kurtz, *notes of the lecture on stochastic analysis*, University of Wisconsin, 2009.

[11] Michael, J., Schucany, W., Haas, R., *Generating random variates using transformations with multiple roots*, The American Statistician 30(2), 88-90, 1976.

[12] P. Protter, *Stochastic Integration and Differential Equations*, Springer-Verlag, Heidelberg, 1990.

[13] Jin Qin, Jerry Lawless, *Empirical likelihood and general estimating equations*, The Annals of Statistics, Vol. 22, N° 1, 300-325, 1994.

[14] J. Rosinski, *Contribution to the discretization of a paper by Barndorff-Nielsen and Shephard*, Journal of the Royal Statistics society B 63, 230-231, 2001.

[15] J. Rosinski, *Tempered Stable Processes, in O.E. Barndorff-Nielsen (Ed.)*, Miniproceding of 2nd MaPhySto Conference on Lévy Processes: Theory and Application, 215-220, 2002.

[16] K.Sato, *Lévy processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge, 1999.

[17] W. Schoutens, *Lévy Processes in Finance: Pricing Financial Derivatives*, John Wiley & Sons, 2003.

[18] D. Stirzaker, *Stochastic Processes and Models*, Oxford University Press, Oxford, 2005.

[19] E. Taufer, N. Leonenko, *Simulation of Lévy-driven Ornstein-Uhlenbeck processes with given marginal distribution*.

[20] Valdivieso, L., Schoutens, W. & Tuerlinckx, *Maximum likelihood estimation in processes of Ornstein-Uhlenbeck type* F. Stat Inference Stoch Process, 2009.